IBM SPSS Statistics 19 Core System 用户指南

Note: Before using this information and the product it supports, read the general information under Notices第 384 页码.

前言

IBM SPSS Statistics

IBM® SPSS® Statistics 是一种用于分析数据的综合系统。SPSS Statistics可以从几乎任何类型的文件中获取数据，然后使用这些数据生成分布和趋势、描述统计以及复杂统计分析的表格式报告、图表和图。

本手册《IBM SPSS Statistics 19 Core System 用户指南》记录了 SPSS Statistics 的图形用户界面。随软件安装的帮助系统中提供了使用附加选项中所包含统计过程的示例。

此外，在菜单和对话框之下，SPSS Statistics 使用命令语言。该系统的一些扩展功能只能通过命令语法来访问。（这些功能在 Student Version 中不提供。）命令语法的详细参考信息，以两种方式提供：集成在完整的“帮助”系统中，在“帮助”菜单的命令语法参考中，也以单独的 PDF 格式文档提供。

IBM SPSS Statistics 选项

以下选项作为完整（非 Student Version）IBM® SPSS® Statistics Core 系统的附加增强功能提供：

Statistics Base 为您提供广泛的统计过程以执行基本分析和报告，其中包括计数、交叉表和描述统计、OLAP 立方和码本报告。它还提供了多种降维、分类和细分方法，例如因子分析、聚类分析、最近邻元素分析和判别函数分析。此外，SPSS Statistics Base 还提供了广泛的平均值比较算法和预测方法，例如 t 检验、方差分析、线性回归和序数回归。

Advanced Statistics 主要提供一些复杂试验和生物医学研究中常用的技术。它包括一些用于一般线性模型 (GLM)、线性混合模型、方差成分分析、对数线性分析、序数回归、保险精算寿命表、Kaplan-Meier 生存分析以及基础和扩展 Cox 回归的过程。

Bootstrap 方法可以导出稳健的标准误估计值，并能为诸如均值、中位数、比例、几率比、相关系数或回归系数等估计值导出置信区间。

Categories 执行最优尺度过程，包括对应分析。

Complex Samples 使得调查、市场、卫生和民意研究者以及使用抽样调查方法的社会学家能够将他们复杂的样本设计并入数据分析中。

Conjoint 提供度量单个产品属性如何影响消费者和市民偏好的现实的方式。使用 Conjoint，您可以轻松地在一组产品属性的环境中度量每个产品属性的折衷效应—如同消费者在进行采购决策时所做的一样。

Custom Tables 可创建各种具有演示质量的表格式报表，包括复杂的行列表和多重响应数据的显示。

Data Preparation 提供数据的快捷可视快照。使用它可以应用标识无效数据值的验证规则。您可以创建标记超出范围的值、缺失值或空值的规则。还可以保存记录个别规则的违反和每个个案的规则违反总数的变量。还提供了可以复制或修改的一组有限的预定义规则。

Decision Trees 创建基于树的分类模型。它将个案分为若干组，或根据自变量（预测变量）的值预测因变量（目标变量）的值。此过程为探索性和证实性分类分析提供验证工具。

Direct Marketing 使组织能够通过专为直销设计的方法确保其营销计划尽可能地发挥效力。

Exact Tests 可在小样本或分布非常不均匀的样本可能导致常规检验不准确的情况下计算统计检验的精确的 p 值。此选项只在 Windows 操作系统中可用。

Forecasting 通过使用多种曲线拟合模型、平滑模型和用于估计自回归函数的方法，执行综合的预测和时间序列分析。

Missing Values 描述了缺失数据的模式、估计均值和其他统计量，并为缺失观察值归因值。

Neural Networks可以通过将产品需求预测为价格函数以及其他变量的函数或根据购买习惯和人口统计特征分类客户来制定经营决策。神经网络是非线性数据建模工具。它们可以用来建立输入与输出之间的复杂关系模型，也可用来查找数据中的模式。

Regression 提供了用于分析那些不能拟合传统线性统计模型的数据的方法。它包括一些用于 probit 分析、logistic 回归、权重估计、两阶段最小平方回归和常规非线性回归的过程。

Amos（**矩™结构的™分析™**）使用结构化方程建模以确认和解释涉及态度、观念和其他驱动行为的因素的概念模型。

关于 SPSS Inc.， IBM 下属公司

SPSS Inc. 是一家 IBM 下属公司，它也是全球领先的预测分析软件和解决方案提供商。该公司拥有全面的产品系列，涵盖数据收集、统计量、建模和部署，通过在业务流程中嵌入分析技术，收集人们的态度与看法，预测未来客户交互结果，然后针对这些深入见解采取相应行动。SPSS Inc. 解决方案着眼于整合分析技术、IT 基础设施和业务流程，以帮助达成整个企业内相互关联的业务目标。全球各地的众多企业、政府和学术机构客户依靠 SPSS Inc. 技术在吸引、留住和发展客户方面取得竞争优势，同时减少欺诈并缓解风险。SPSS Inc. 在 2009 年 10 月被 IBM 并购。有关更多信息，请访问 http://www.spss.com。

技术支持

我们提供有“技术支持”以维护客户。客户可就 SPSS Inc. 产品使用或某一受支持硬件环境的安装帮助寻求技术支持。要获得“技术支持”，请访问 SPSS Inc. 网站 http://support.spss.com，或通过网站 http://support.spss.com/default.asp?refpage=contactus.asp 找到当地办事处。在请求协助时，请准备好您和您组织的 ID 以及支持协议。

客户服务

如果对发货或帐户存在任何问题，请联系您当地的办事处，联系方式列在 Web 站点中，网址为 http://www.spss.com/worldwide。请先准备好您的序列号以供识别。

培训讲座

SPSS Inc. 提供公开的以及现场的培训讲座。所有讲座都是以实践小组为特色的。讲座将定期在各大城市开展。关于这些讲座的更多信息，请联系您本地的办事处，联系方式列在 Web 站点上，网址为 http://www.spss.com/worldwide。

附加出版物

SPSS Statistics：数据分析指南、SPSS Statistics：Statistical Procedures Companion 和 SPSS Statistics：Advanced Statistical Procedures Companion（由 Marija Norušis 编写，并已由 Prentice Hall 出版）作为建议的补充材料提供。这些出版物涵盖 SPSS Statistics Base 模块、Advanced Statistics 模块和 回归模块中的统计过程。无论您是刚开始从事数据分析工作，还是已准备好使用高级应用程序，这些书籍都将帮助您最有效地利用在 IBM® SPSS® Statistics 产品中找到的功能。有关其他信息，包括出版物的内容和示例章节，请参阅作者的网站： http://www.norusis.com

内容

4 分布式分析模式 60

5 数据编辑器 66

8 数据转换 119

12 模型 236

13 使用命令语法 249

14 图表工具的概述 262

15 使用预测模型对数据评分 272

16 实用程序 280

17 选项 289

章

1

概述

19 版本中的新增功能

线性模型。 线性模型根据目标与一个或多个预测变量间的线性关系来预测连续目标。线性模型相对简单，用于评分的数学公式也易于解释。这些模型的属性比较好理解，与同一数据集上的其他模型类型（如神经网络或决策树）相比能够非常快速构建。此功能在 Statistics Base 附加模块中可用。

广义线性混合模型。 广义线性混合模型扩展了线性模型，使得：目标通过指定的关联函数与因子和协变量线性相关；目标可以有非正态分布；观测可能相关。广义线性混合模型涵盖了从简单线性回归到复杂的非正态纵向数据多变量模型的各种模型。此功能在 Advanced Statistics 附加模块中可用。

轻量表。 轻量表可以比全功能枢轴表更快呈现。尽管缺乏枢轴表的编辑功能，但它们很容易转换为所有编辑功能可用的枢轴表。

评分向导。 新的评分向导使您可以轻松应用预测模型对数据评分，并且评分不再需要 IBM® SPSS® Statistics 服务器。

改进的默认测量级别。 对于从外部来源读取的数据以及在会话中创建的新变量，确定默认测量级别的方法已有所改进，以评估多个条件而不仅仅是唯一值的数量。由于测量级别会影响许多过程的结果，因此正确的测量级别指派通常是非常重要的。

“智能”输出。 直销附加模块的过程现在提供“智能”输出：帮助您评估结果的简单、非技术性说明。

语法编辑器增强。 您现在可以将编辑器窗格拆分为上下两个窗格。您可以增加或减少语法块的缩进，或者采用与所粘贴语法类似的格式自动缩进选定内容。新的工具栏按钮允许您取消注释先前被注释的文本，并且新的选项设置允许您将语法粘贴在光标位置。您现在还可以导航到下一个或上一个语法错误（例如，不匹配的引号），这样很容易在运行语法之前找到这些错误。

salesforce.com 的数据库驱动程序。 salesforce.com 的数据库驱动程序允许分析人员访问 salesforce.com 中的数据，就像您在访问 SQL 数据库中的数据一样。分析人员现在可以连接到 salesforce.com，提取相关数据并执行分析。

编译转换。 在您使用编译转换时，转换命令（如 COMPUTE 和 RECODE）在运行期间被编译成机器代码，以提升那些包含大量个案的数据集的转换性能。此功能需要 SPSS Statistics 服务器。

Statistics portal。Statistics portal 是一个基于 Web 的 IBM® SPSS® Collaboration and Deployment Services 用户界面，允许这些用户使用 SPSS Statistics 引擎的强大功能来分析他们的数据。用户从在 SPSS Statistics 中编写的自定义用户界面（使用自定义对话框生成器）上运行分析，并存储在他们的 IBM SPSS Collaboration and Deployment Services Repository 中。有关 Statistics portal 的自定义用户界面的作者的功能增强包括：在相邻分析之间接受为活动数据集指定的过滤器；在 CROSSTABS、OLAP CUBES 和 CTABLES 生成的表中隐藏较小计数；并在 CROSSTABS 交叉表中将一组行和列维度显示为表层。

Windows

IBM® SPSS® Statistics 中有一些不同类型的窗口：

数据编辑器。数据编辑器显示数据文件的内容。您可以用数据编辑器创建新的数据文件或修改现有的数据文件。如果打开了多个数据文件，则每个数据文件都有一个单独的数据编辑器窗口。

查看器。所有的统计结果、表格和图表都显示在“查看器”中。您可以编辑输出并进行保存，便于以后使用。“查看器”窗口在您第一次运行生成输出的过程时自动打开。

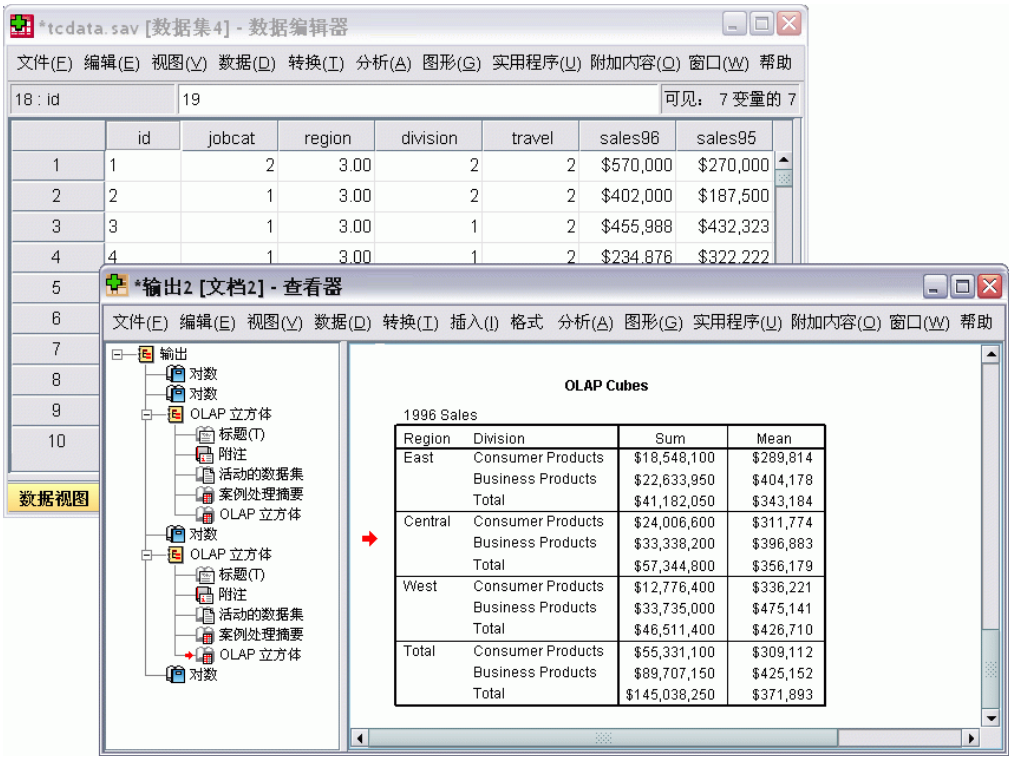
枢轴表编辑器。使用枢轴表编辑器可以通过多种方法修改显示在枢轴表中的输出。您可以编辑文本，交换行中和列中的数据，添加颜色，创建多维表，以及选择隐藏和显示结果。

图表编辑器。您可以修改图表窗口中的高分辨率图表以及绘图。您可以更改颜色，选择不同类型的字体或大小，切换水平轴和垂直轴，旋转三维散点图，甚至更改图表类型。

文本输出编辑器。没有显示在枢轴表中的文本输出可以用文本输出编辑器进行修改。可以编辑输出并更改字体特征（类型、样式、颜色、大小）。

语法编辑器。可以将对话框中的选择内容粘贴到语法窗口，在语法窗口中您的选择显示为命令语法格式。然后可以编辑命令语法，以使用不能通过对话框使用的特殊功能。您可以将这些命令保存在文件中，便于在以后的会话中使用。

图片 1-1
数据编辑器和查看器



指定的窗口和活动窗口

如果您打开了多个“查看器”窗口，输出会转到**指定的**“查看器”窗口。如果打开了多个语法编辑器窗口，命令语法会粘贴到指定的语法编辑器窗口中。指定的窗口在标题栏中用加号图标表示。您可以随时更改指定的窗口。

指定的窗口不应与**活动**窗口相混淆，活动窗口是当前选中的窗口。如果您有重叠的窗口，则活动窗口显示在最前面。如果您打开一个窗口，该窗口就自动成为活动窗口和指定的窗口。

更改指定的窗口

▶ 使您要指定的窗口成为活动窗口（单击窗口中的任意位置）。

▶ 单击工具栏上的“指定窗口”按钮（加号图标）。

或

▶ 从菜单中选择：

实用程序 > 指定窗口

注意：对于数据编辑器窗口，活动的数据编辑器窗口确定在后续计算或分析中使用的数据集。没有“指定的”数据编辑器窗口。

状态栏

每个 IBM® SPSS® Statistics 窗口底部的状态栏提供以下信息：

命令状态。对于您运行的每个过程或命令，都会有一个个案计数器指示到目前为止已经处理的个案数。对于需要迭代处理的统计过程，将显示迭代的次数。

过滤状态。如果您已选择了要分析的随机样本或个案子集，**过滤范围**消息表明当前正在进行某种类型的个案过滤，并且不是数据文件中的所有个案都包括在分析中。

权重状态。加权范围消息表明正在使用权重变量对要分析的个案进行加权。

拆分文件状态。分割文件范围消息表明数据文件已按照一个或多个分组变量的值分割成单个组以进行分析。

对话框

大多数菜单选择都会打开对话框。可以使用对话框选择要分析的变量和分析选项。

统计过程和图表的对话框一般有两个基本组件：

源变量列表。活动数据集中的变量的列表。只有选中的过程所允许的变量类型才会显示在源列表中。短字符串和长字符串变量的使用在许多过程中受到限制。

目标变量列表。指示您已选择用于分析的变量的一个或多个列表，如因变量和自变量列表。

对话框列表中的变量名和变量标签

可以在对话框列表中显示变量名或变量标签，并且可以控制变量列表中的变量排序顺序。要控制源列表中变量的默认显示属性，请从“编辑”菜单中选择**选项**。

您也可以在对话框中更改变量列表显示属性。更改显示属性的方法因对话框而异：

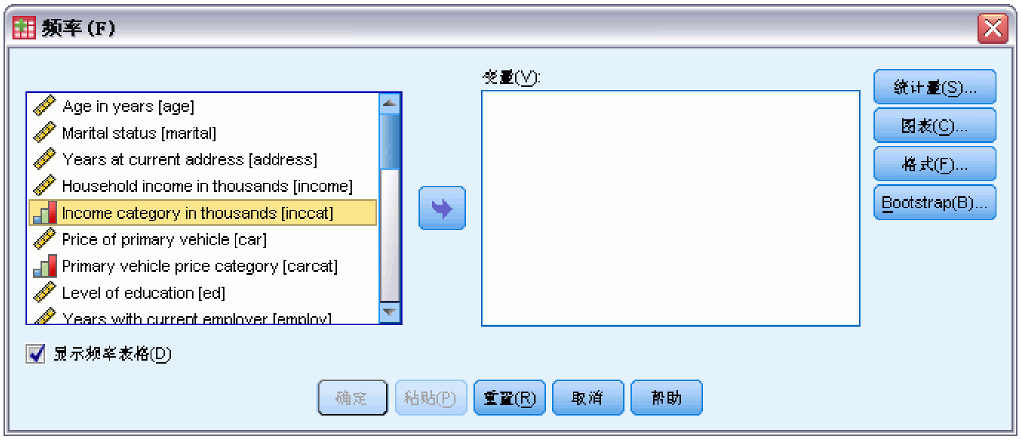
- 如果对话框在源变量列表上方提供有排序和显示控件，则使用这些控件更改显示属性。
- 如果对话框在源变量列表上方未包含排序控件，请右键单击源列表中的任何变量，然后从上下文菜单中选择显示属性。

可以显示变量名或变量标签（对于未定义标签的变量，显示变量名），并且可以按文件顺序、字母顺序或测量级别对源列表进行排序。（在源变量列表上方具有排序控件的对话框中，默认选择**无**将按文件顺序对列表排序。）

“调整大小”对话框

您可以像窗口那样，单击并拖动外部边框或角落来调整对话框大小。例如，如果使对话框变宽，变量列表也将变得更宽。

图片 1-2
“调整大小”对话框



对话框控件

大多数对话框中都有 5 个标准控件：

确定或**运行**。运行过程。选择变量并选择任何其他指定后，请单击**确定**运行过程并关闭对话框。某些对话框包含**运行**按钮，而不是“确定”按钮。

粘贴。从对话框选择生成命令语法，并将语法粘贴到语法窗口中。然后可以使用对话框中没有的其他功能定制命令。

重置。取消选择选中的变量列表中的任何变量，将对话框和所有子对话框中的所有指定重置为默认状态。

取消。取消对话框设置中自上次打开对话框以来所做的任何更改，并关闭对话框。在一次会话中，对话框设置是不变的。对话框将保留上次的指定信息，直到覆盖它们。

帮助。提供上下文相关的帮助。该控件将打开一个“帮助”窗口，其中包含关于当前对话框的信息。

选择变量

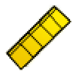
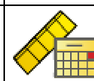
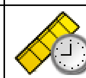
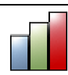
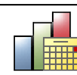
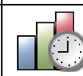
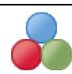
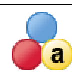
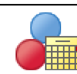
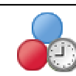
要选择单个变量，只需在源变量列表中选择该变量，然后将其拖放到目标变量列表。还可以使用箭头按钮将变量从源列表移动到目标列表。如果只有一个目标变量列表，则可以双击单个变量，将其从源列表移动到目标列表。

也可以选择多个变量：

- 要选择变量列表中分在一组中的多个变量，请单击第一个变量，然后按住 Shift 再单击组中的最后一个变量。
- 要选择变量列表上没有分在一组中的多个变量，请单击第一个变量，然后按住 Ctrl 再单击下一个变量，依此类推（Macintosh：Command-click）。

数据类型、测量级别和变量列表图标

显示在对话框列表中的变量旁边的图标提供有关变量类型和测量级别的信息。

测量级别	数据类型			
	数值	字符串	日期	时间
尺度（连续）		n/a		
有序				
名义				

- 有关测量级别的更多信息，请参见变量测量级别第 69 页码。
- 有关数值、字符串、日期和时间数据类型的更多信息，请参见变量类型第 70 页码。

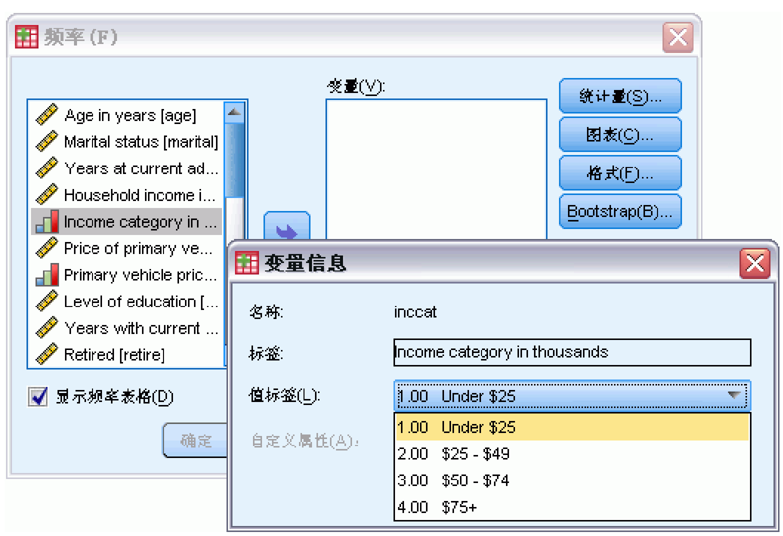
获得关于对话框中的变量的信息

许多对话框提供有查看有关在变量列表 所显示变量的更多信息的功能。

▶ 右键单击源变量列表或目标变量列表中的变量。

▶ 选择变量信息。

图片 1-3
变量信息



数据分析中的基本步骤

用 IBM® SPSS® Statistics 分析数据非常简单。您需要做的只是：

将数据输入 SPSS Statistics。您可以打开先前保存的 SPSS Statistics 数据文件；读取电子表格、数据库或文本数据文件；或者将数据直接输入数据编辑器。

选择一个过程。从菜单中选择一个过程来计算统计量或创建图表。

选择要分析的变量。数据文件中的变量显示在该过程的对话框中。

运行过程并查看结果。结果显示在查看器中。

统计辅导

如果您不熟悉 IBM® SPSS® Statistics 或其中可用的统计过程，“统计指导”可以帮助您入门。它用简单的问题、非技术性的语言和可视的示例提示您，帮助您选择最适合您的数据的基本统计功能和图表绘制功能。

要使用“统计指导”，请从任意 SPSS Statistics 窗口的菜单中选择：

帮助 > 统计辅导

“统计指导”仅涵盖选定的过程子集。它旨在对许多基本的、常用的统计技巧提供帮助。

了解更多信息

要获得基础知识的综合概述，请参见联机教程。从任意 IBM® SPSS® Statistics 菜单中选择：

帮助 > 教程

章

2

获得帮助

帮助以多种不同形式提供：

“帮助”菜单。大多数窗口中的“帮助”菜单可以访问主帮助系统、教程和技术参考材料。

- **主题。** 可以访问“目录”、“索引”和“搜索”选项卡，使用这些选项卡可以查找特定帮助主题。
- **教程。**有关如何使用众多基本功能的分步图解说明。不必通读整个教程。可以选择要查看的主题，以任何顺序跳过和查看主题，使用索引或目录查找特定主题。
- **个案研究。**如何创建各种类型的统计分析以及如何解释结果的实践示例。同时还提供示例所使用的样本数据文件，使您能够通过对示例的研究，实际了解结果是如何生成的。可以从目录选择要了解的特定过程，或在索引中搜索相关主题。
- **统计指导。**类似于向导的方法，指导您完成查找要使用的过程。进行一系列选择后，“统计指导”将打开用于符合所选标准的统计、报告或绘图过程的对话框。
- **命令语法参考。** 命令语法的详细参考信息，以两种方式提供：集成在完整的“帮助”系统中，在“帮助”菜单的命令语法参考中，以单独的 PDF 格式文档提供。
- **统计算法。** 以两种形式提供用于大多数统计过程的算法：与整个帮助系统集成，或以单独的 PDF 文档，可以从手册 CD 中获得。有关指向帮助系统中的特定算法的链接，请在“帮助”菜单中选择“**算法**”。

上下文相关的帮助。在用户界面中的许多地方都可以获得上下文相关的帮助。

- **对话框“帮助”按钮。**大多数对话框都有“帮助”按钮，点击该按钮可直接进入该对话框的“帮助”主题。“帮助”主题提供一般信息和相关主题的链接。
- **枢轴表上下文菜单帮助。**右键单击浏览器中已激活的枢轴表里的项，并在上下文菜单中选择**这是什么？**以显示这些项的定义。
- **命令语法。**在命令语法窗口中，将光标放在命令的语法块中的任意位置，然后按键盘上的 F1。此时将显示该命令的完整命令语法图表。可从相关主题列表中的链接中以及“帮助目录”选项卡中获得完整的命令语法文档。

其他资源

技术支持网站。在以下网站可以找到许多常见问题的解答：http://support.spss.com。（技术支持网站需要登录 ID 和密码。上面列出的 URL 提供了有关如何获得 ID 和密码的信息。）

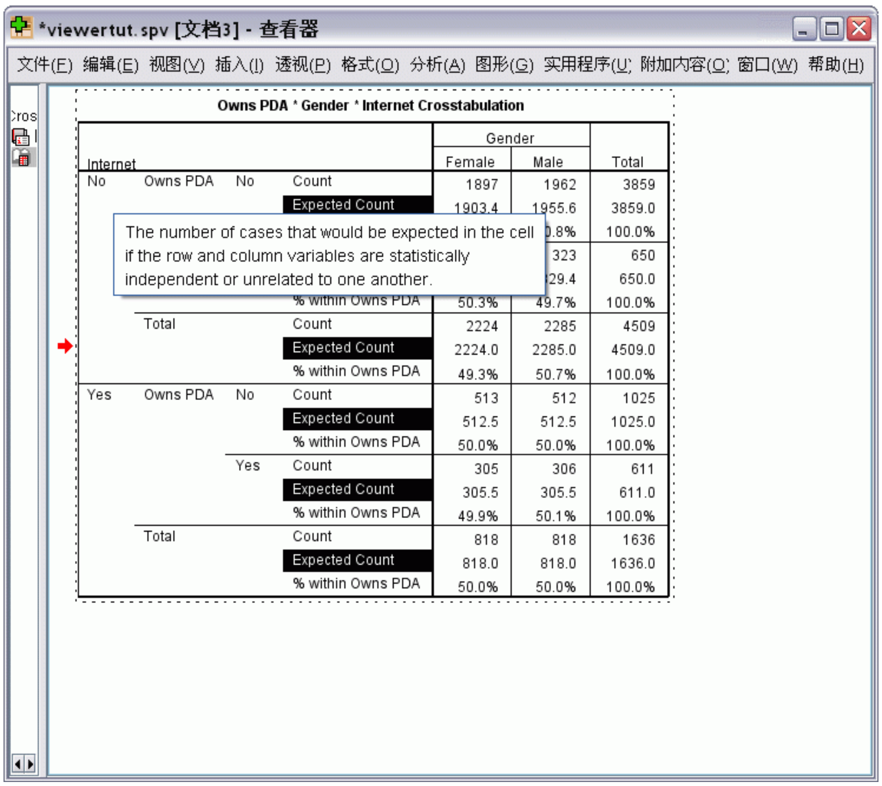
Developer Central。Developer Central 拥有适合所有级别用户和应用程序开发人员的资源。下载实用程序、图形示例、新统计模块和文章。请访问 Developer Central，网址为http://www.spss.com/devcentral。

获得输出项帮助

要在浏览器的枢轴表输出中查看某项的定义：

- ▶ 双击激活枢轴表。
- ▶ 右键单击想要得到解释的项。
- ▶ 从上下文菜单中，选择**这是什么？**。

在弹出窗口中将显示该项的定义。



图片 2-1
单击鼠标右键可出现已激活的枢轴表词汇表帮助

章

3

数据文件

数据文件有多种格式，而本软件被设计为可以处理其中的许多格式，包括：

- 用 Excel 和 Lotus 创建的电子表格
- 来自许多数据库源（包括 Oracle、SQLServer、Access、dBASE 和其他）的数据库表
- 以 Tab 分隔的和其他类型的简单文本文件
- 在其他操作系统上创建的 IBM® SPSS® Statistics 格式的数据文件
- SYSTAT 数据文件
- SAS 数据文件
- Stata 数据文件

打开数据文件

除了以 IBM® SPSS® Statistics 格式保存的文件以外，还可以打开 Excel、SAS、Stata、制表符分隔文件和其他文件，而无需将文件转换为中间格式或输入数据定义信息。

- 打开数据文件会使其成为活动数据集。如果已经打开了一个或多个数据文件，则它们将保持打开状态，并可在以后的会话中使用。单击“数据编辑器”窗口中的任意位置会使打开的数据文件成为活动数据集。
- 在分布式分析中，使用远程服务器处理命令和运行过程的模式、可用的数据文件、文件夹和驱动器取决于远程服务器上可用的内容。当前服务器名称在对话框的顶部指明。除非将驱动器指定为共享设备，或者将包含数据文件的文件夹指定为共享文件夹，否则将不能访问本地计算机上的数据文件。

打开数据文件

▶ 从菜单中选择：

文件 > 打开 > 数据...

▶ 在“打开数据”对话框中，选择要打开的文件。

▶ 单击**打开**。

根据需要，您可以：

- **根据观察值，最小化字符串宽度**将每个字符串变量宽度自动设置为该变量的最长观察值。在 Unicode 模式中读取代码页数据文件时特别有用。
- 从电子表格文件的第一行读取变量名。

- 指定电子表格文件中要读取的单元范围。
- 指定 Excel 文件中要读取的工作表（Excel 95 或更高版本）。

有关从数据库中读取数据的信息，请参见读取数据库文件第 13 页码。有关从文本数据文件中读取数据的信息，请参见文本向导第 27 页码。

数据文件类型

SPSS Statistics。 打开以 IBM® SPSS® Statistics 格式保存的数据文件以及 DOS 产品 SPSS/PC+。

SPSS/PC+。 打开 SPSS/PC+ 数据文件。此选项只在 Windows 操作系统上可用。

SYSTAT。 打开 SYSTAT 数据文件。

SPSS Statistics 便携。 打开以可移植格式保存的数据文件。以便携格式保存文件比以 SPSS Statistics 格式保存文件所耗费的时间要长得多。

Excel。 打开 Excel 文件。

Lotus 1-2-3。 打开以 1-2-3 格式（Lotus R3.0、2.0 或 1A）保存的数据文件。

SYLK。 打开以 SYLK（符号链接）格式保存的数据文件，这是某些电子表格应用程序使用的格式。

dBASE。 打开 dBASE 格式文件（dBASE IV、dBASE III 或 III PLUS 或者 dBASE II）。每个个案均是一条记录。当您以这种格式保存文件时，变量和值标签以及缺失值的指定会丢失。

SAS。 SAS 版本 6 - 9 和 SAS 传输文件。使用命令语法，您还可以从 SAS 格式目录文件中读取值标签。

Stata。 Stata 版 4 - 8。

打开文件选项

读取变量名称。 对于电子表格，您可以从文件的第一行或定义范围的第一行读取变量名。按需要转换值，以创建有效的变量名，例如将空格转换为下划线。

工作表。 Excel 95 或更高版本的文件可以包含多个工作表。缺省情况下，数据编辑器读取第一张工作表。要读取其它工作表，请从下拉列表中选择工作表。

范围。 对于电子表格数据文件，您还可以读取某个单元范围。请使用与在电子表格应用程序中相同的方法指定单元范围。

读取 Excel 95 或更高版本的文件

以下规则适用于读取 Excel 95 或更高版本的文件：

数据类型和宽度。每一列都是一个变量。每个变量的数据类型和宽度都由 Excel 文件中的数据类型和宽度决定。如果该列包含多个数据类型（例如：日期和数字），数据类型就设置为字符串，所有的值都读取为有效的字符串值。

空白单元格。对于数值变量，空白单元格会转换为系统缺失的值，用句点表示。对于字符串变量，空格是有效的字符串值，空白单元格被视为有效的字符串值。

变量名称。如果将 Excel 文件的第一行（或者指定范围的第一行）读取为变量名称，则不符合变量命名规则的值会转换为有效的变量名称，原始名称用作变量标签。如果不从 Excel 文件读取变量名称，则会指定缺省的变量名称。

读取旧 Excel 文件和其他电子表格

以下规则适用于读取早于 Excel 95 的 Excel 文件以及其他电子表格数据：

数据类型和宽度。 每个变量的数据类型和宽度由列中第一个数据单元的列宽和数据类型确定。其他类型的值会转换为系统缺失值。如果列中的第一个数据单元是空白的，则使用该电子表格的全局缺省数据类型（通常为数值）。

空白单元格。对于数值变量，空白单元格会转换为系统缺失的值，用句点表示。对于字符串变量，空格是有效的字符串值，空白单元格被视为有效的字符串值。

变量名称。 如果不从电子表格读取变量名，则列字母（A、B、C...）用于 Excel 和 Lotus 文件的变量名。对于以 R1C1 显示格式保存的 SYLK 文件和 Excel 文件，本软件使用以字母 C 开头的列号作为变量名（C1、C2、C3 等）。

读取 dBASE 文件

数据库文件在逻辑上与 IBM® SPSS® Statistics 数据文件非常相似。以下一般规则适用于 dBASE 文件：

- 字段名称会转换为有效的变量名。
- 用于 dBASE 字段名称的冒号会转换为下划线。
- 包含标记为要删除但未实际清除的记录。本软件创建一个新的字符串变量 D_R，该变量对标记为要删除的个案包含一个星号。

读取 Stata 文件

以下一般规则适用于 Stata 数据文件：

- **变量名称。**Stata 变量名以区分大小写的形式转换为 IBM® SPSS® Statistics 变量名。通过附加下划线和顺序字母（_A、_B、_C、...、_Z、_AA、_AB、...等等），将只有大小写不同的 Stata 变量名转换为有效的变量名。
- **变量标签。**Stata 变量标签转换为 SPSS Statistics 变量标签。
- **值标签。**Stata 值标签转换为 SPSS Statistics 值标签，但归为“扩展”缺失值的 Stata 值标签除外。
- **缺失值。**Stata“扩展”缺失值转换为系统缺失值。
- **日期转换。** Stata 日期格式值转换为 SPSS Statistics DATE 格式 (d-m-y) 值。Stata“时间序列”日期格式值（周数、月数、季度数等）转换为简单数值 (F) 格式，同时保留原始内部整数值，即从 1960 年开始算起的周数、月数、季度数等等。

读取数据库文件

只要有某种数据库格式的数据库驱动程序，就可以读取该数据库格式的数据。用本地分析模式时，本地计算机上必须装有所需的驱动程序。用分布式分析模式时（IBM® SPSS® Statistics Server 提供），远程服务器上必须安装这些驱动程序。

注意：如果您正在运行 SPSS Statistics 的 Windows 64 位版本，则无法读取 Excel、Access 或 dBASE 数据库源，即使它们出现在可用数据库源列表中。这些产品的 32 位 ODBC 驱动程序不兼容。

读取数据库文件

▶ 从菜单中选择：
文件 > 打开数据库(B) > 新建查询...

▶ 选择数据源。

▶ 如果需要（取决于数据源），可选择数据库文件和/或输入登录名、密码和其他信息。

▶ 选择表和字段。对于 OLE DB 数据源（仅在 Windows 操作系统上可用）只能选择一个表。

▶ 指定表之间的关系。

▶ 或者：

- 为数据指定任何选择条件。
- 添加一个提示，供用户输入信息以创建参数查询。
- 运行构建的查询之前请先保存。

编辑已保存的数据库查询

▶ 从菜单中选择：
文件 > 打开数据库(B) > 编辑查询...

▶ 选择要编辑的查询文件 (*.spq)。

▶ 请按照创建新查询的说明操作。

使用已保存的查询读取数据库文件

▶ 从菜单中选择：
文件 > 打开数据库(B) > 运行查询...

▶ 选择要运行的查询文件 (*.spq)。

▶ 如果需要（取决于数据库文件），输入登录名和密码。

▶ 如果查询包含嵌入的提示，则根据需要输入其他信息（例如，要检索销售数据的季度）。

选择数据源

使用“数据库向导”的第一个屏幕选择要读取的数据源类型。

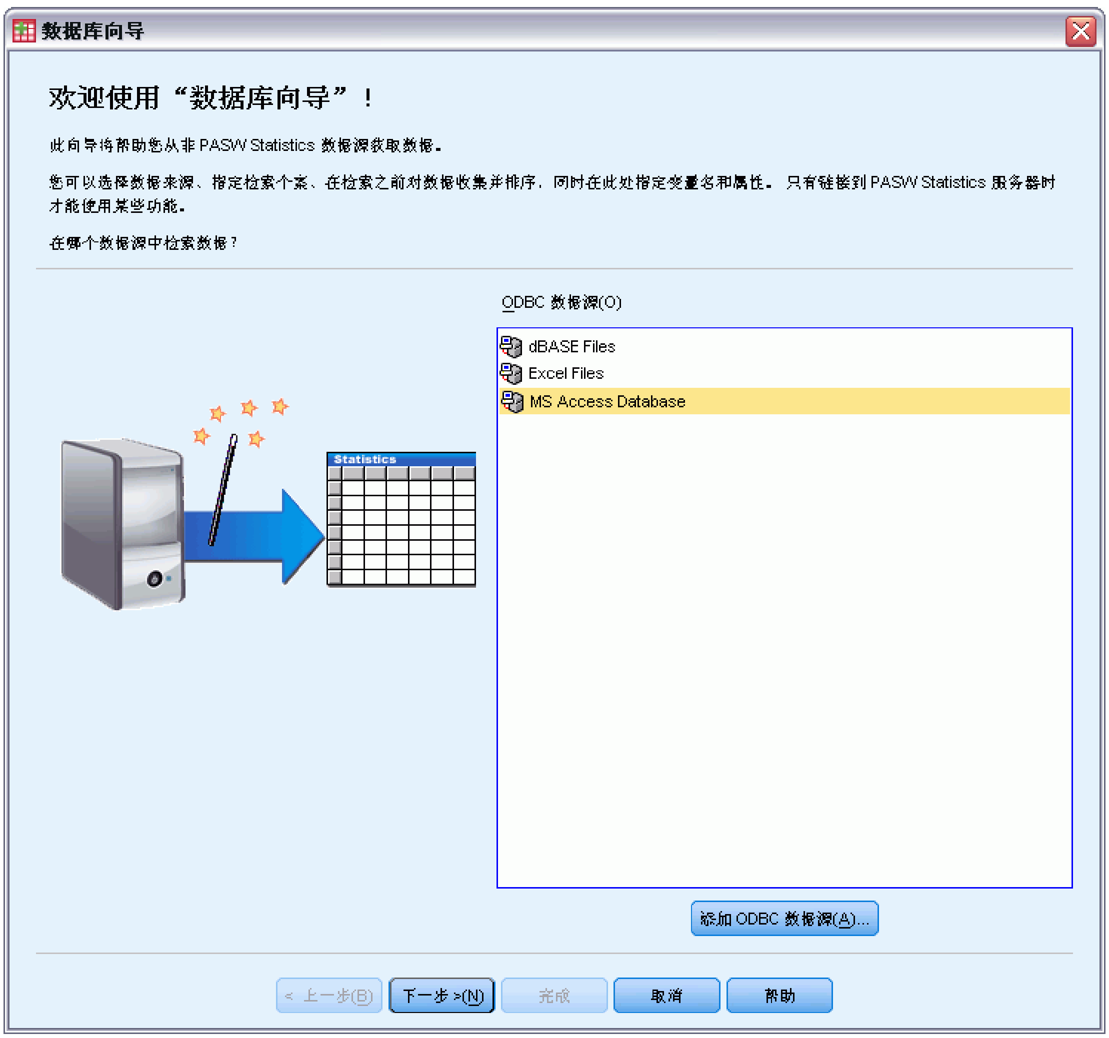
ODBC 数据源(O)

如果没有配置任何 ODBC 数据源，或者要添加新的数据源，请单击**添加 ODBC 数据源**。

- 在 Linux 操作系统中，该按钮不可用。在 odbc.ini 中指定 ODBC 数据源，并且 ODBCINI 环境变量必须设定为该文件的位置。有关更多信息，请参见数据库驱动程序文档。
- 用分布式分析模式时（IBM® SPSS® Statistics Server 提供），该按钮不可用。要用分布式分析模式添加数据源，请咨询系统管理员。

ODBC 数据源由两条基本信息组成：要用来访问数据的驱动程序和要访问的数据库的位置。要指定数据源，必须装有适当的驱动程序。针对不同数据库格式的驱动程序可以从 http://www.spss.com/drivers 获得。

图片 3-1
数据库向导



OLE DB 数据源

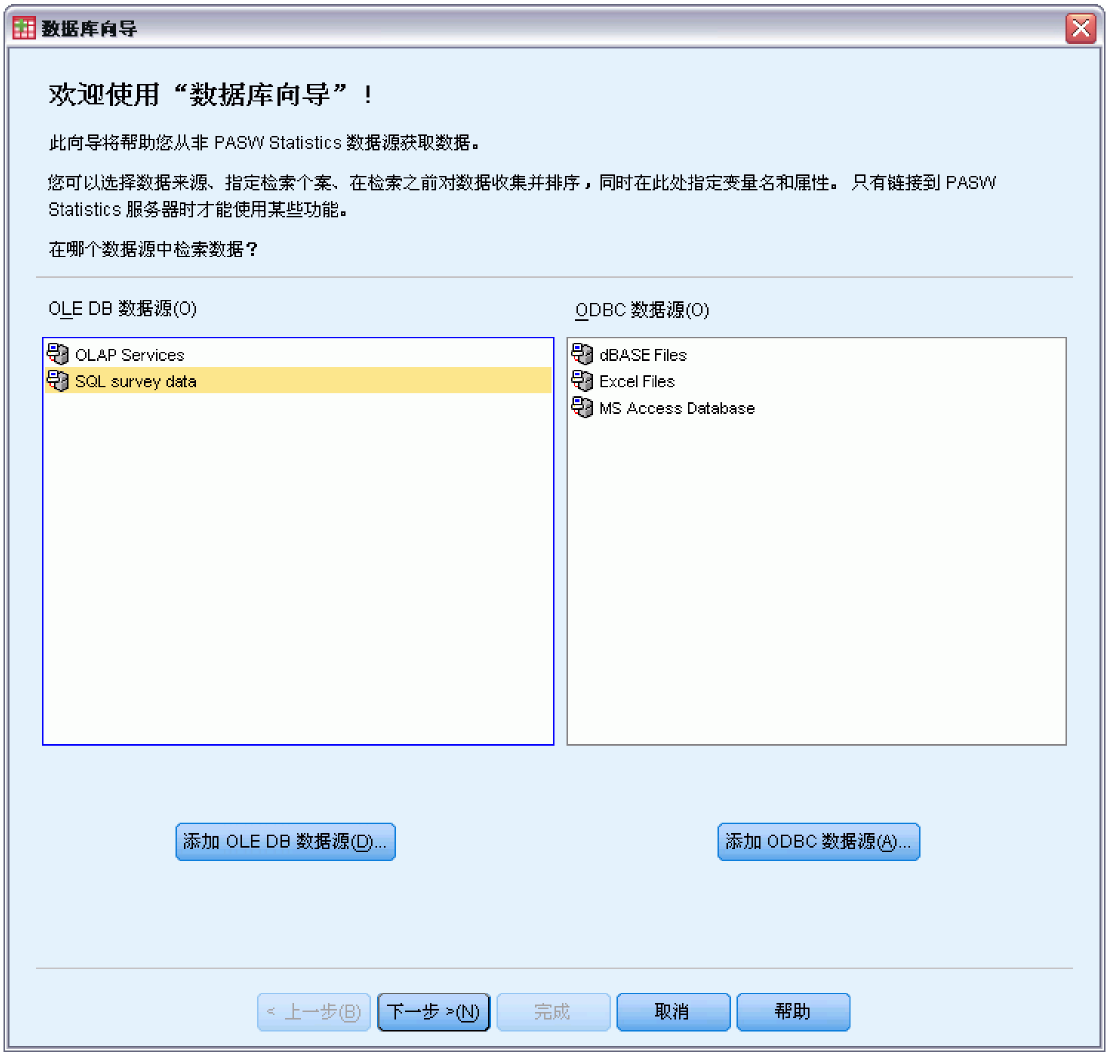
要访问 OLE DB 数据源（在 Microsoft Windows 操作系统上可用），必须安装有以下内容：

- .NET Framework。要获得最新版本的 .NET Framework，请转至 http://www.microsoft.com/net。
- IBM® SPSS® Data Collection Survey Reporter Developer Kit。可以从安装介质上安装与此发行版本兼容的某个版本。如果您在使用 IBM® SPSS® Statistics Developer，则可以从 www.spss.com/statistics (http://www.spss.com/statistics/) 的“下载”选项卡上下载兼容的版本。

以下限制适用于 OLE DB 数据源：

- OLE DB 数据源不支持表连接。一次只能读取一个表。
- 您只能以本地分析模式添加 OLE DB 数据源。要在 Windows 服务器上以分布式分析模式添加 OLE DB 数据源，请咨询系统管理员。
- 使用分布式分析模式时（SPSS Statistics Server 提供），OLE DB 数据源只能在 Windows 服务器上使用，并且该服务器上必须安装有 .NET 和 SPSS Survey Reporter Developer Kit。

图片 3-2
使用数据库向导访问 OLE DB 数据源



添加 OLE DB 数据源：

- 单击**添加 OLE DB 数据源**。
- 在“数据链接属性”对话框中，单击**提供程序**选项卡，然后选择 OLE DB 提供程序。
- 单击**下一步**，或单击**连接**选项卡。
- 可通过输入目录位置和数据库名称，或者单击按钮浏览到数据库来选择数据库。（还可能需要用户名和密码。）
- 输入所有必需信息后，单击**确定**。（您可通过单击**测试连接**按钮来确保指定的数据库可用。）
- 为数据库连接信息输入名称。（该名称将显示在可用的 OLE DB 数据源列表中。）

图片 3-3
“将 OLE DB 连接信息另存为”对话框



▶ 单击**确定**。

此操作将返回到“数据库向导”的第一个屏幕，您可以在该屏幕上从 OLE DB 数据源列表中选择已保存的名称，然后继续执行向导的下一步。

删除 OLE DB 数据源

要从 OLE DB 数据源列表中删除数据源名称，可在下列位置删除包含该数据源名称的 UDL 文件：

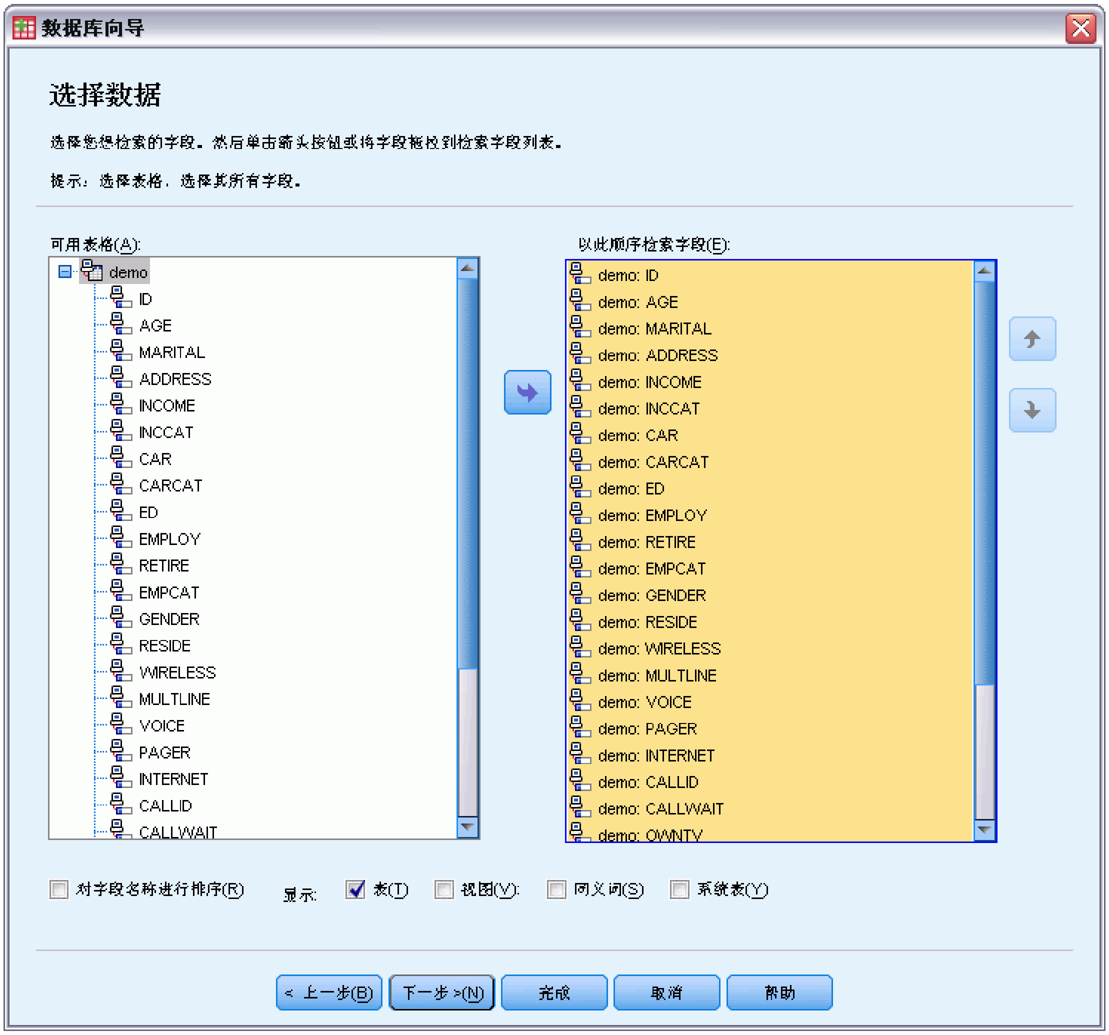
[驱动器]:\Documents and Settings\[user login]\Local Settings\Application Data\SPSS\UDL

选择数据字段

“选择数据”步骤控制将读取哪些表和字段。数据库字段（列）读取为变量。

如果在表中选择了任何字段，则在以下“数据库向导”窗口中所有的字段都可见，但只有在这一步中选中的字段才作为变量导入。这使您可以创建表连接，并使用未导入的字段指定条件。

图片 3-4
数据库向导，选择数据



显示字段名称。要列出表中的字段，请单击表名称左边的加号 (+)。要隐藏字段，请单击表名称左边的减号 (–)。

要添加字段。双击“可用的表”列表中的任何字段，或者将其拖到“按此顺序检索字段”列表中。在字段列表中拖放字段可以对其重新排序。

删除字段。双击“按此顺序检索字段”列表中的任何字段，或者将其拖到“可用的表”列表。

将字段名称排序。如果选中该复选框，“数据库向导”就会按照字母顺序显示可用的字段。

默认情况下，可用表的列表只显示标准数据库表。可以控制列表中显示的项的类型：

- **表。**标准数据库表。

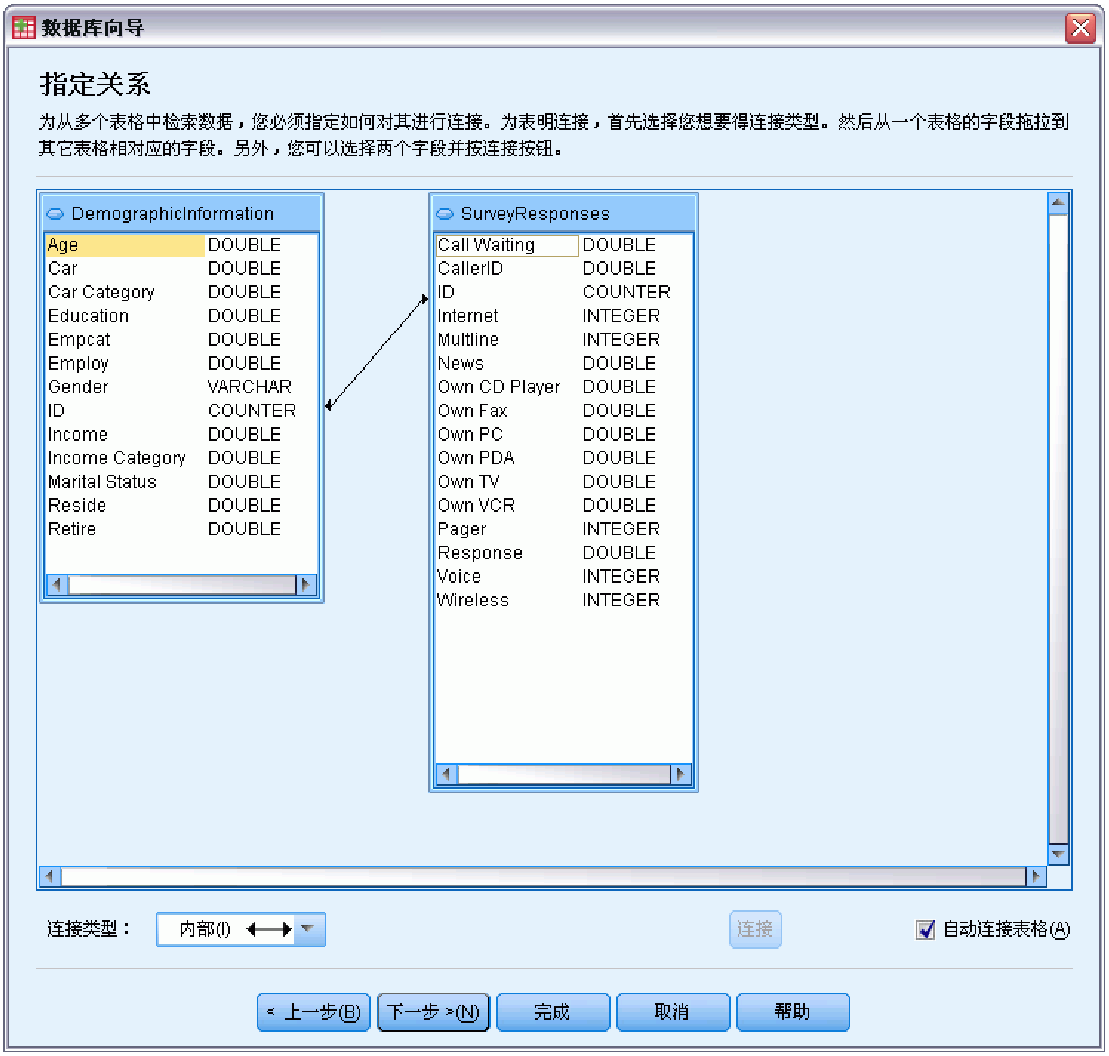
- **视图。**视图是由查询定义的虚拟的或动态的“表”。视图中可以包含基于其他字段值计算得出的多个表和/或字段的连接。
- **同义词。**同义词是表或视图的别名，通常在查询中定义。
- **系统表。**系统表定义数据库的属性。在某些情况下，标准数据库表可能会被分类成系统表，并且仅在选择了该选项后才会显示。通常只有数据库管理员才有访问真正的系统表的权限。

注意：对于 OLE DB 数据源（仅在 Windows 操作系统上可用），只能从单个表中选择字段。OLE DB 数据源不支持多个表连接。

创建表之间的关系

“指定关系”步骤使您可以定义 ODBC 数据源的表之间的关系。如果选择的字段来自一个以上的表，则必须定义至少一个连接。

图片 3-5
数据库向导，指定关系



建立关系。要创建关系，请将任意表上的字段拖到要连接的字段上。“数据库向导”将在两个字段之间画一条**连接线**，表明它们的关系。这些字段的数据类型必须相同。

自动连接表。可尝试按照主/外键或匹配的字段名和数据类型来自动连接表。

连接类型。如果驱动程序支持外部连接，则可以指定内部连接、左边外连接或右边外连接。

- **Inner joins.** 内部连接仅包括相关字段相等的行。在此示例中，将包括两个表中具有匹配的 ID 值的所有行。
- **外部连接**。除了内部连接的一对一匹配外，还可以使用外部连接通过一对多匹配方案来合并表。例如，您可以将其中只包含少量代表数据值和关联描述性标签的记录的表，和包含上百个或上千个代表调查响应者的记录的表中的值相匹配。左边外连接包括左边的表中的所有记录，而仅包括右边的表中相关字段相等的记录。在右边外连接中，连接从右边的表导入所有记录，而仅从左边的表导入相关字段相等的记录。

限制检索的个案

“限制检索的个案”步骤使您可以指定选择个案（行）的子集的条件。限制个案通常包括用条件填充条件网格。条件由两个表达式以及它们之间的某种关系组成。该表达式返回每个个案的 true、false 或 missing 值。

- 如果结果是 true，则选中该个案。
- 如果结果是 false 或 missing，则不选中该个案。
- 大多数条件使用六个关系运算符中的一个或多个（<、>、<=、>=、= 和 <>）。
- 表达式可以包括字段名、常数、算术运算符、数字和其他函数以及逻辑变量。可以将不打算导入的字段用作变量。

图片 3-6
数据库向导，限制检索的个案



要建立条件，至少需要两个表达式和一种连接表达式的关系。

▶ 要建立表达式，请选择下列一种方法：

- 在“表达式”单元格中，键入字段名、常数、算术运算符、数字和其他函数或逻辑变量。
- 双击“字段”列表中的一个字段。
- 将字段从“字段”列表中拖到“表达式”单元格。
- 从任何活动的“表达式”单元格的下拉菜单中选择一个字段。

▶ 要选择关系运算符（如 = 或 >），请将光标放在“关系”单元格上，然后键入运算符或从下拉菜单中进行选择。

如果 SQL 包含具有个案选择表达式的 WHERE 子句，则表达式中的日期和时间需要以特殊方式指定（包括示例中显示的花括号）：

- 应使用一般形式 {d 'yyyy-mm-dd'} 来指定日期文本。
- 应使用一般格式 {t 'hh:mm:ss'} 来指定时间文本。
- 应使用一般格式 {ts 'yyyy-mm-dd hh:mm:ss'} 来指定日期/时间文本（时间戳）。
- 整个日期和/或时间值都必须用单引号括起。年份必须以四位数的形式表示；日期和时间的值的每个部分都必须包含两位数。例如，2005 年 1 月 1 日上午 1:05 应表示为以下形式：

  {ts '2005-01-01 01:05:00'}

函数。可选择内置算术、逻辑、字符串、日期和时间 SQL 函数。可将函数从列表中拖到表达式中，或者输入任何有效的 SQL 函数。关于有效的 SQL 函数，请参阅您的数据库文档。可从以下位置获得标准函数的列表：

http://msdn2.microsoft.com/en-us/library/ms711813.aspx

使用随机抽样。该选项从数据源选择个案的随机样本。对于大数据源，您可能需要将个案数限制为小的、具有代表性的样本，这可以显著减少其运行程序所需的时间。本机随机抽样（如果对该数据源可用）速度比 IBM® SPSS® Statistics 随机抽样要快，因为 SPSS Statistics 随机抽样必须读取整个数据源才能抽取随机样本。

- **近似**. 生成近似于指定个案百分比的随机样本。由于此例程为每个个案作出独立的伪随机决策，因此选定个案的百分比只能近似于指定的百分比。数据文件中的个案越多，选定个案的百分比与指定百分比就越接近。
- **准确**. 从指定的个案总数中选择指定数目的个案作为随机样本。如果指定的个案总数超过数据文件中的个案总数，则样本将按比例包含比请求数目少的个案。

注意：如果使用随机抽样，则分类汇总（SPSS Statistics Server 中的分布式模式提供）不可用。

输入值提示。可以在查询中嵌入一个提示来创建**参数查询**。当用户运行查询时，系统将要求用户输入信息（根据此处指定的信息）。如果要查看同一数据的不同视图，则可能要进行这一操作。例如，您可能想要运行相同的查询来查看不同财政季度的销售数据。

▶ 将光标放在任何“表达式”单元格中，然后单击**输入值提示**来创建提示。

创建参数查询

使用“输入值提示”步骤来创建一个对话框，在每次有人运行查询时请求用户提供信息。如果要用不同的条件查询相同的数据源，则该功能将很有用。

图片 3-7
值的提示



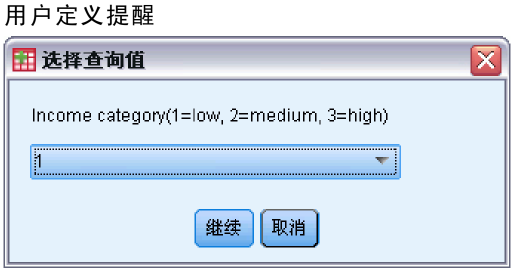
要建立提示，请输入提示字符串和默认值。每次用户运行查询时，该提示字符串都会显示。该字符串应指定要输入的信息类型。如果用户不从列表中进行选择，那么该字符串应给出有关如何设置输入格式的提示。示例如下：输入季度（Q1、Q2、Q3...）。

允许用户从列表中选择值。如果选中该复选框，您可以限制用户选择您放在此处的值。确保使用回车分隔值。

数据类型。在此处选择数据类型（数字、字符串或日期）。

最后的结果看起来如下所示：

图片 3-8
用户定义提醒



分类汇总数据

如果处于分布式模式中，并已连接到远程服务器（IBM® SPSS® Statistics Server 提供），那么可以先对数据进行分类汇总，然后再将其读入 IBM® SPSS® Statistics 中。

图片 3-9
数据库向导，汇总数据



还可以在将数据读到 SPSS Statistics 中之后再对其进行分类汇总，但对于大数据源来说，预先进行分类汇总可以节省时间。

▶ 要创建分类汇总数据，请选择一个或多个定义如何分组个案的分隔变量。

▶ 选择一个或多个分类汇总变量。

▶ 为每个分类汇总变量选择一个分类汇总函数。

▶ 或者，创建包含每个分类组中的个案数的变量。

注意：如果使用 SPSS Statistics 随机抽样，分类汇总将不可用。

定义变量

变量名和标签。完整的数据库字段（列）名用作变量标签。除非您修改变量名，否则“数据库向导”将按照以下两种方法中的一种将变量名指定给数据库的每一列：

- 如果数据库字段的名称是有效的、唯一的变量名，则该名称将用作变量名。
- 如果数据库字段的名称不是有效的、唯一的变量名，那么将自动生成一个新的唯一名称。

单击任何单元格来编辑变量名。

将字符串转换为数字值。如果要自动将字符串变量转换为数值型变量，请针对该字符串变量选择**重新编码为数值型**框。字符串值按照原始值的字母顺序转换为连续的整数值。原始值保留为新的变量的值标签。

变量宽度字符串字段的宽度。该选项控制变量宽度字符串值的宽度。默认情况下，宽度为 255 个字节，并且只读取前 255 个字节（通常指单字节语言中的 255 个字符）。宽度至多可以为 32,767 个字节。尽管您可能不需要截断字符串值，但也不要指定不必要的长值，这会导致处理效率很低。

根据观察值，最小化字符串宽度。自动将每个字符串变量的宽度设置为最长观察值。

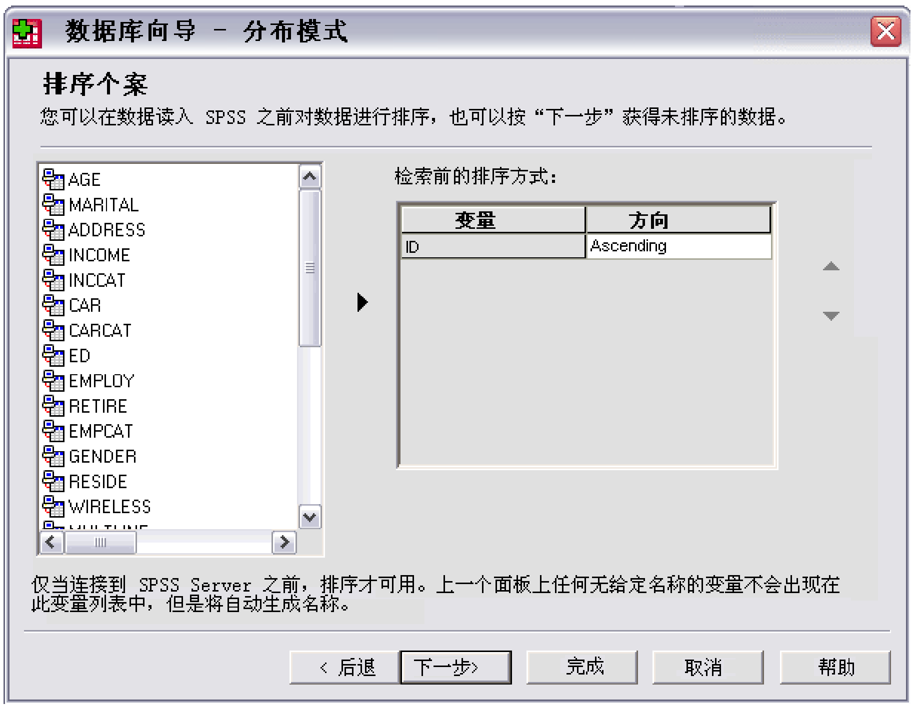
图片 3-10
数据库向导，定义变量



排序个案

如果处于分布式模式中，并已连接到远程服务器（IBM® SPSS® Statistics Server 提供），那么可以先对数据进行分类，然后再将其读入 IBM® SPSS® Statistics 中。

图片 3-11
数据库向导，对个案进行排序



还可以在将数据读到 SPSS Statistics 中之后再对其进行排序，但对于大数据源来说，预先进行排序可以节省时间。

结果

“结果”步骤显示查询的 SQL Select 语句。

- 可以在运行查询前编辑该 SQL Select 语句，但是如果单击**上一步**按钮在前面的步骤中进行更改，那么对 Select 语句所做的更改将丢失。
- 要保存查询以供将来使用，可使用**将查询保存到文件**区段。
- 要将完整的 GET DATA 语法粘贴到语法窗口，请选择**将其粘贴到语法编辑器以供将来修改**。从“结果”窗口复制和粘贴 Select 语句将不会粘贴所需的命令语法。

注意：粘贴的语法在由向导生成的每一行 SQL 中的结束引号之前包含一个空格。这些空格不是多余的。在处理命令时，SQL 语句的所有行将以一种文本形式合并在一起。如果没有空格，在一行的最后一个字符和下一行的第一个字符之间就不会存在空格。

图片 3-12
数据库向导，结果面板



文本向导

文本向导可以读取多种格式的文本数据文件：

- 制表符分隔文件
- 空格分隔文件
- 逗号分隔文件
- 固定字段格式的文件

对于分隔的文件，也可以将其他字符指定为值之间的分隔符，并且可以指定多个分隔符。

读取文本数据文件

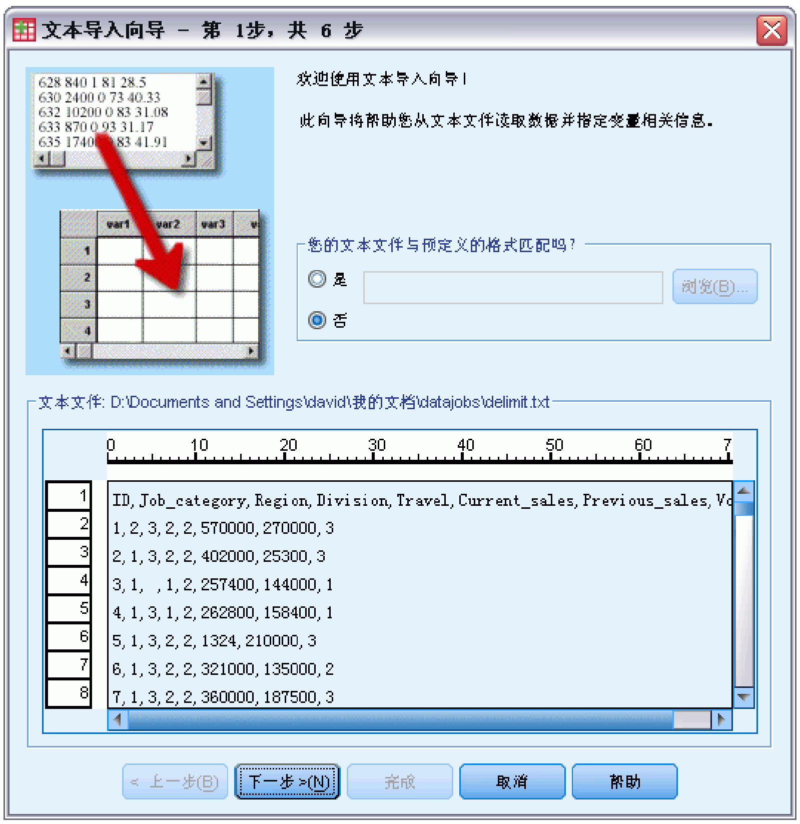
▶ 从菜单中选择：
文件 > 读取文本数据...

▶ 在“打开数据”对话框中选择文本文件。

▶ 按照文本向导中的步骤来定义如何读取数据文件。

文本向导：第 1 步

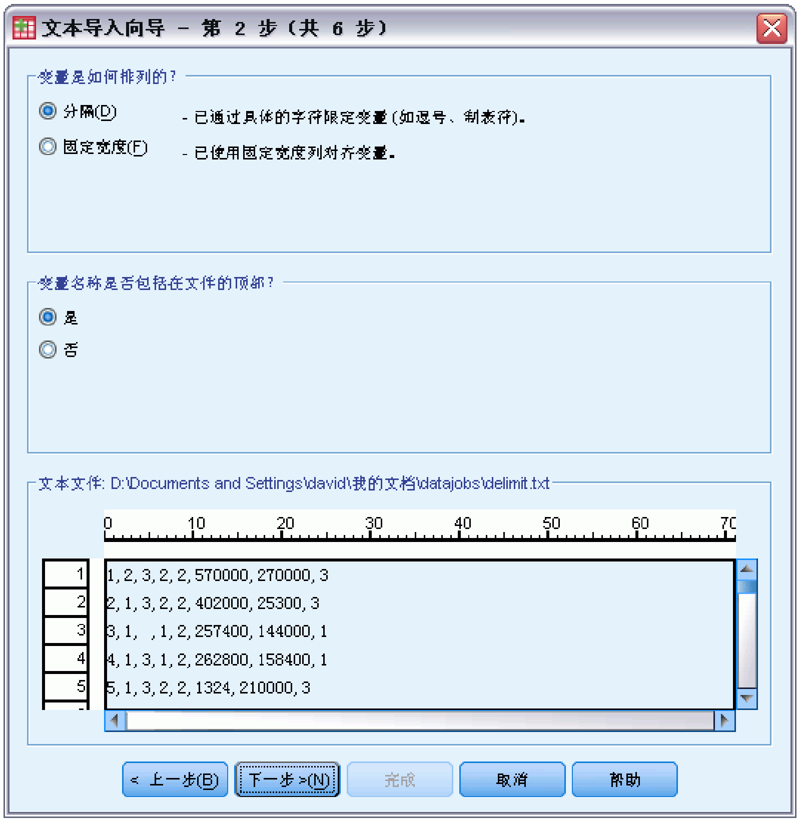
图片 3-13
文本向导：第 1 步



文本文件显示在一个预览窗口中。可以应用预定义的格式（以前在文本向导中保存的），或者按照文本向导中的步骤来指定如何读取数据。

文本向导：第 2 步

图片 3-14
文本向导：第 2 步



此步骤提供变量的信息。变量类似于数据库中的字段。例如，问卷中的每一项都是一个变量。

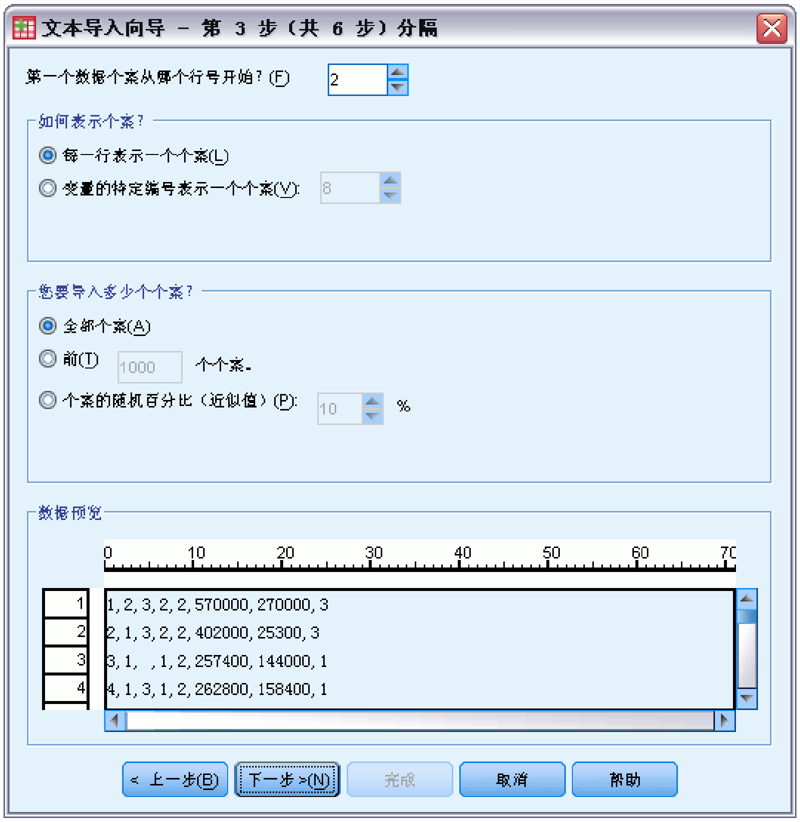
变量是如何排列的？要正确读取您的数据，文本向导需要知道如何确定一个变量的数据值结束且下一个变量的数据值开始的位置。变量的安排定义用于将一个变量与另一个变量区分开来的方法。

■ **分隔。**使用空格、逗号、制表符和其他字符分隔变量。变量为每个个案按照同样的顺序进行记录，但不一定在相同的列位置。

■ **固定宽度。**对于数据文件中的每个个案，每个变量都记录在同一个记录（行）上的相同列位置。变量之间不需要分隔符。实际上，在许多由计算机程序生成的文本数据文件中，数据值可能显示为连在一起，互相之间甚至没有用空格分隔开。列位置确定要读取的是哪个变量。

变量名称是否包括在文件的顶部？如果数据文件的第一行包含每个变量的描述性标签，则可以将这些标签用作变量名。不符合变量命名规则的值会转换为有效的变量名。

文本向导：第 3 步（分隔的文件）



图片 3-15
文本向导：第 3 步（对于分隔的文件）

这一步提供有关个案的信息。个案类似于数据库中的记录。例如，问卷的每个回答者都是一个个案。

第一个数据个案从哪个行号开始？(F) 表示包含数据值的数据文件的第一行。如果数据文件的顶行包含描述性标签或者包含不代表数据值的其他文本，这就不是第 1 行。

如何表示个案？控制文本向导如何确定每个个案结束、下一个个案开始的位置。

- **每一行表示一个个案。**每一行仅包含一个个案。每个个案通常包含在一个单行中，即使这一行对于有大量变量的数据文件会很长。如果不是所有的行都包含相同数量的数据值，则每个个案的变量数由数据值的个数最多的行决定。对于数据值较少的个案，多出来的变量指定为缺失值。
- **变量的特定编号表示一个个案。**每个实例的指定变量数告诉文本向导在哪里停止读取某个个案，并开始读取下一个个案。同一行可以包含多个个案，个案可以在一行的中间开始，并在下一行继续。文本向导按照读取的值的数量确定每个个案的结束，不管有多少行。每个个案必须包含所有变量的数据值（或者由分隔符表示的缺失值），否则数据文件将无法正确读取。

您要导入多少个个案？您可以导入数据文件中的所有个案，可以导入前 n 个个案（n 是您指定的数字），也可以随机导入指定百分比的样本。因为随机抽样程序对每个个案都作出独立的假随机决策，所以选定的个案的百分比可能只与指定的百分比相近。数据文件中的个案越多，选定个案的百分比与指定百分比就越接近。

文本向导：第 3 步（固定宽度的文件）

图片 3-16
文本向导：第 3 步（对于固定宽度的文件）



这一步提供有关个案的信息。个案类似于数据库中的记录。例如，问卷中的每个回答者都是一个个案。

第一个数据个案从哪个行号开始？(F)表示包含数据值的数据文件的第一行。如果数据文件的顶行包含描述性标签或者包含不代表数据值的其他文本，这就不是第 1 行。

多少行表示一个个案？(L)控制文本向导如何确定每个个案结束、下一个个案开始的位置。每个变量由其个案内的行数及其列位置定义。需要指定每个个案的行数，才能正确读取数据。

您要导入多少个个案？您可以导入数据文件中的所有个案，可以导入前 n 个个案（n 是您指定的数字），也可以随机导入指定百分比的样本。因为随机抽样程序对每个个案都作出独立的假随机决策，所以选定的个案的百分比可能只与指定的百分比相近。数据文件中的个案越多，选定个案的百分比与指定百分比就越接近。

文本向导：第 4 步（分隔的文件）

图片 3-17
文本向导：第 4 步（对于分隔的文件）



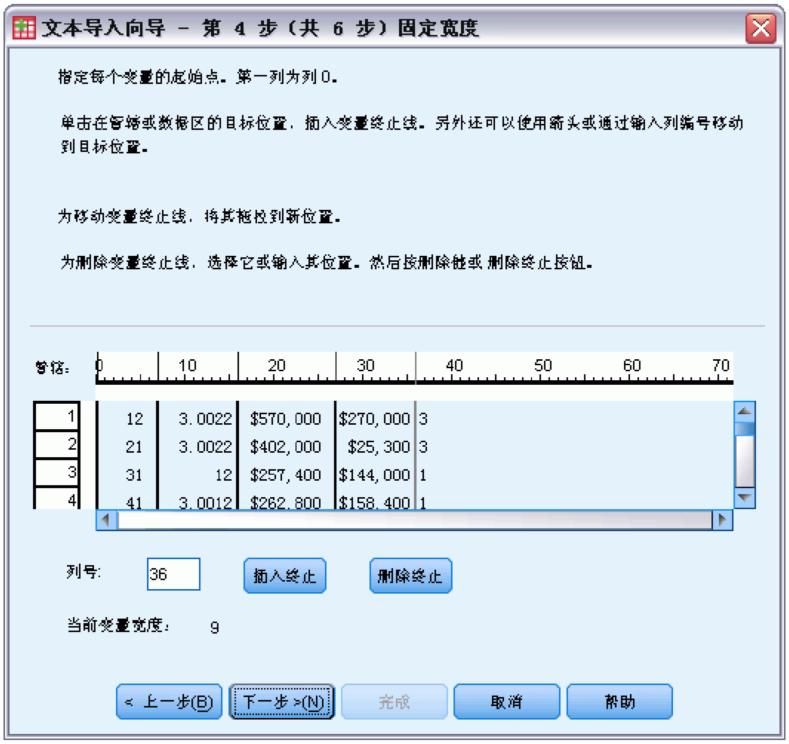
此步骤显示文本向导对于如何读取数据文件的最佳猜测，并使您可以修改文本向导从数据文件读取变量的方式。

变量之间有哪些分隔符？表示隔开数据的值的字符或符号。可以选择空格、逗号、分号、制表符或其他字符的任意组合。中间没有插入数据值的多个连续的分隔符被视为缺失值。

文本限定符是什么？用于包括包含分隔符字符的值的字符。例如，如果逗号是分隔符，包含逗号的值将读取错误，除非有文本限定符封装了该值，使值中的逗号不会被当作值之间的分隔符。从 Excel 导出的 CSV 格式的数据文件使用双引号（“）用作文本限定符。文本限定符出现在值的开头和结尾，封装了整个值。

文本向导：第 4 步（固定宽度的文件）

图片 3-18
文本向导：第 4 步（对于固定宽度的文件）



此步骤显示文本向导对于如何读取数据文件的最佳猜测，并使您可以修改文本向导从数据文件读取变量的方式。预览窗口中的垂直线表示文本向导当前认为每个变量在文件中开始的位置。

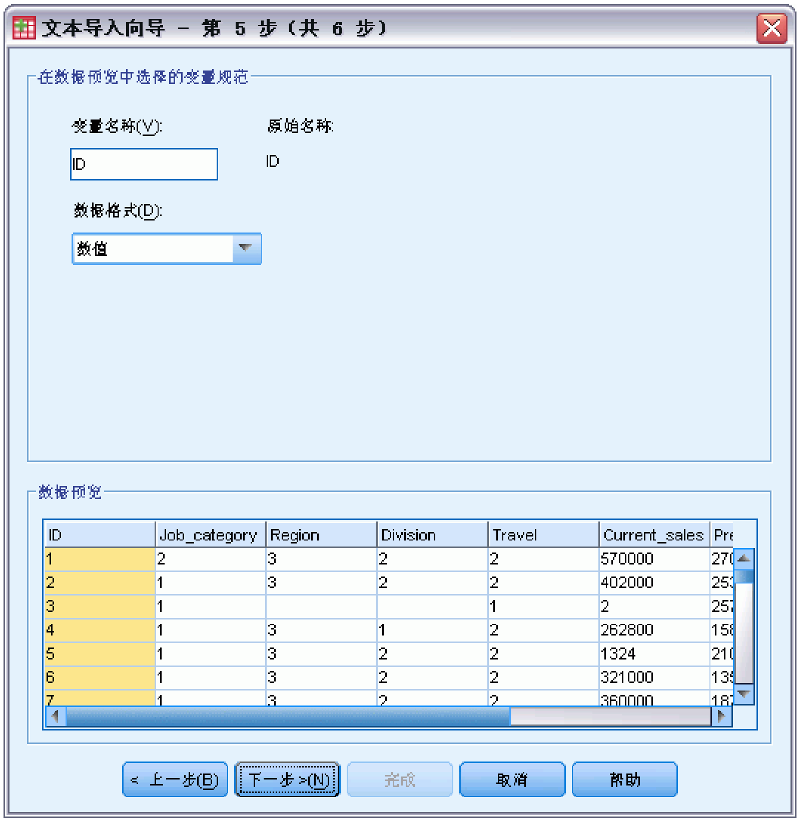
必要时插入、移动和删除变量换行符以分隔变量。如果每个个案使用多行，则数据将按每个个案一行的方式显示，后续行附加在行的末尾。

注意：

对于计算机生成的数据文件，其所生成的一连串连续的数据值没有插入空格或其他明显特征，这样就很难确定每个变量开始的位置。这样的数据文件通常依赖于数据定义文件或其他一些指定每个变量的行和列位置的书面说明。

文本向导：第 5 步

图片 3-19
文本向导：第 5 步



此步骤控制文本向导用于读取每个变量的变量名和数据格式，并制最终数据文件中将包括哪些变量。

变量名称。可以用自己的变量名覆盖缺省的变量名。如果从数据文件读取变量名，文本向导将自动修改不符合变量命名规则的变量名。在预览窗口中选择一个变量，然后输入变量名。

数据格式。在预览窗口选择一个变量，然后从下拉列表选择一种格式。按住 Shift 单击可以选择多个相邻的变量，或者按住 Ctrl 单击可以选择多个不相邻的变量。

缺省格式由前 250 行中的数据值确定。如果在前 250 行中有多种格式（例如，数值、日期、字符串等），则缺省格式设置为字符串。

文本向导格式选项

用文本向导读取变量的格式选项包括：

不导入。省略在导入的数据文件中选择的变量。

数值。有效值包括数字、前导加号或减号以及小数指示符。

字符串。有效值包括几乎任何键盘字符和嵌入的空格。对于分隔文件，可以指定值的字符数量，最多可以指定为 32,767 个。缺省情况下，文本向导将字符数量设置为选定的变量在文件前 250 行中遇到过的最长的字符串值。对于固定宽度的文件，字符串值的字符的数量由步骤 4 中的变量换行符的位置定义。

日期/时间。有效值包括常用的日期格式 dd-mm-yyyy、mm/dd/yyyy、dd.mm.yyyy、yyyy/mm/dd、hh:mm:ss 以及其他各种日期和时间格式。月份可以用数字、罗马数字或三个字母的缩写形式表示，也可以使用全拼的格式。从列表中选择一个日期格式。

美元。有效值为数字，前导美元符号是可选的，作为千位分隔符的逗号也是可选的。

逗号。有效值包括将句点用作小数指示符和将逗号用作千位分隔符的数字。

点。有效值包括将逗号用作小数指示符和将句点用作千位分隔符的数字。

注意：包含对选定的格式无效的字符的值将视为缺失值。包含任何指定分隔符的值将视为多个值。

文本向导：第 6 步

图片 3-20
文本向导：第 6 步



这是文本向导的最后一步。可以将您的规格保存在文件中，以便在导入类似的文本数据文件时使用。也可以将文本向导生成的语法粘贴到语法窗口。然后就可以定制和 / 或保存语法，以便用于其他对话或生产作业中。

在本地缓存数据. 数据高速缓存是数据文件的完整副本，它存储在临时磁盘空间中。高速缓存数据文件可以改进性能。

读取 IBM SPSS Data Collection 数据

在 Microsoft Windows 操作系统上，您可以从 IBM® SPSS® Data Collection 产品读取数据。(注意：此功能只在安装在 Microsoft Windows 操作系统上的 IBM® SPSS® Statistics 中可用。

要读取 Data Collection 数据源，必须安装以下项目：

- .NET Framework。要获得最新版本的 .NET Framework，请转至 http://www.microsoft.com/net。
- IBM® SPSS® Data Collection Survey Reporter Developer Kit。有关获取 SPSS Survey Reporter Developer Kit 兼容版本的信息，请转至 support.spss.com (http://support.spss.com)。

只能以本地分析模式读取 Data Collection 数据源。此功能不能用于使用 SPSS Statistics Server 的分布式分析模式。

从 Data Collection 数据源读取数据：

▶ 从打开的任意 SPSS Statistics 窗口中的菜单选择：
文件 > 打开 Data Collection 数据

▶ 在“数据链接属性”的“连接”选项卡上，指定元数据文件、个案数据类型和个案数据文件。

▶ 单击**确定**。

▶ 在“Data Collection 数据导入”对话框中，选择要包括的变量并选择个案选择标准。

▶ 单击**确定**以读取数据。

“数据链接属性”：“连接”选项卡

要读取 IBM® SPSS® Data Collection 数据源，需要指定：

元数据位置。 包含问卷定义信息的元数据文档文件 (.mdd)。

个案数据类型。 个案数据文件的格式。可用格式包括：

- Quancept 数据文件 (DRS)。Quancept .drs、.drz 或 .dru 文件中的个案数据。
- Quanvert 数据库。Quanvert 数据库中的个案数据。
- Data Collection 数据库 (MS SQL Server)。SQL Server 中关系数据库中的个案数据。
- Data Collection XML 数据文件。XML 文件中的个案数据。

个案数据位置。包含个案数据的文件。此文件的格式必须与所选个案数据类型一致。

注意：“连接”选项卡上的其他设置或其他“数据链接属性”选项卡上的任何设置对将 Data Collection 数据读取到 IBM® SPSS® Statistics 中的影响程度尚未知，因此建议不要更改其中任何设置。

“选择变量”选项卡

可以选择要读取的变量的子集。缺省情况下，将显示并选中数据源中的所有标准变量。

- **显示系统变量。** 显示任何“系统”变量，包括指示访问状态（进行中、已完成、完成日期等）的变量。随后可以选择要包含的任何系统变量。缺省情况下将排除所有系统变量。
- **显示代码变量。**显示任何表示代码（用于对分类变量未确定的“其他”响应）的变量。随后可以选择要包含的任何代码变量。缺省情况下将排除所有代码变量。
- **显示 SourceFile 变量。** 显示任何包含已扫描响应的映像文件名的变量。随后可以选择要包含的任何 SourceFile 变量。缺省情况下将排除所有 SourceFile 变量。

“个案选择”选项卡

对于包含系统变量的 IBM® SPSS® Data Collection 数据源，可以基于一定数目的系统变量标准选择个案。您不必在要读取的变量的列表中包括相应的系统变量，但是在源数据中必须存在必要的系统变量才能应用选择标准。如果源数据中不存在必要的系统变量，则忽略相应的选择标准。

数据收集状态。可以选择响应者数据、测试数据或两者都选择。还可以基于以下访问状态参数的任意组合来选择个案：

- 成功完成
- 活动/进行中
- 超时
- 被脚本中止
- 被响应中止
- 访问系统关闭
- 信号（已由脚本中的信号语句终止）

数据收集结束日期。 可以基于完成数据收集的日期选择个案。

- **起始日期。**包括其数据收集在指定日期或该日期之后完成的个案。
- **结束日期。** 包括其数据收集在指定的日期之前完成的个案。这不包括其数据收集在结束日期完成的个案。
- 如果同时指定起始日期和结束日期，则这会定义从起始日期到（但不包括）结束日期的完成日期范围。

文件信息

数据文件包含的内容远不止是原始数据。它还包含所有变量定义信息，包括：

- 变量名

- 变量格式
- 描述性的变量标签和值标签

这些信息存储在数据文件的字典部分。数据编辑器提供一种查看变量定义信息的方法。还可以显示活动数据集或任何其他数据文件的完整字典信息。

显示数据文件信息

▶ 从数据编辑器窗口的菜单中选择：
文件 > 显示数据文件信息

▶ 对于当前打开的数据文件，请选择**工作文件**。

▶ 对于其他数据文件，选择**外部文件**，然后选择数据文件。
浏览器中会显示数据文件信息。

保存数据文件

除了以 IBM® SPSS® Statistics 格式保存数据文件之外，还可以用多种外部格式保存数据，这些格式包括：

- Excel 和其他电子表格格式
- Tab 分隔和 CSV 文本文件
- SAS
- Stata
- 数据库表

保存已修改的数据文件

▶ 使数据编辑器成为活动窗口（单击窗口的任何部位即可使其成为活动窗口）。

▶ 从菜单中选择：
文件 > 保存

已修改的数据文件被保存，并覆盖文件的上一版本。

注意：16.0 版之前的 IBM® SPSS® Statistics 版本无法读取以 Unicode 格式保存的数据文件。要保存早期发行版可以读取的 Unicode 数据文件，在代码页中打开文件并将其重新保存。将根据当前区域编码格式保存该文件。如果包含当前区域无法识别的字符，可能会丢失一些数据。有关切换 Unicode 模式和代码页模式的信息，请参见一般选项第 290 页码。

以外部格式保存数据文件

▶ 使数据编辑器成为活动窗口（单击窗口的任何部位即可使其成为活动窗口）。

▶ 从菜单中选择：
文件 > 另存为...

- 从下拉列表中选择文件类型。
- 为新数据文件输入文件名。

要将变量名写入电子表格或 tab 分隔数据文件的第一行，请执行以下操作：

- 单击“将数据保存为”对话框中的**将变量名称写入电子表格**。

要以 Excel 文件保存值标签而非数据值，请执行以下操作：

- 单击“将数据保存为”对话框中的**在已定义值标签时保存值标签而不是保存数据值**。

要将值标签保存到 SAS 语法文件（仅当选择了 SAS 文件类型时是活动的），请执行以下操作：

- 单击“将数据保存为”对话框中的**将值标签保存到 .sas 文件**。

有关将数据导出到数据库表的信息，请参见 导出到数据库第 44 页码。

有关导出数据以用于 IBM® SPSS® Data Collection 应用程序的信息，请参见 导出到 IBM SPSS Data Collection第 55 页码。

保存数据：数据文件类型

您可以用以下格式保存数据：

SPSS Statistics (*.sav)。 IBM® SPSS® Statistics 格式。

- 7.5 版之前的软件无法读取以 SPSS Statistics 格式保存的数据文件。16.0 版之前的 SPSS Statistics 版本无法读取以 Unicode 格式保存的数据文件
- 在版本 10.x 或 11.x 中使用有变量名称长于八字节的数据文件时，将使用变量名称的唯一的八字节版本，—但是在发行版 12.0 或更高版本中将保留原变量名称。在 10.0 之前的发行版中，在您保存数据文件时原来的长变量名会丢失。
- 在 13.0 版之前使用字符串变量长于 255 字节的数据文件时，会将这些字符串变量分解为多个 255 字节的字符串变量。

7.0 版 (*.sav)。 7.0 版格式。7.0 版和较早版本的 Windows 版可以读取以 7.0 版格式保存的数据文件，但是不包括已定义的多重响应集或 Data Entry for Windows 信息。

SPSS/PC+ (*.sys)。 SPSS/PC+ 格式。如果数据文件包含的变量多于 500 个，将仅保存前 500 个。对于具有多个已定义用户缺失值的变量，将把其他的用户缺失值记录到第一个已定义用户缺失值中。此格式只在 Windows 操作系统上可用。

SPSS Statistics 便携 (*.por)。 可移植格式，其他版本的SPSS Statistics以及其他操作系统上的版本都可以读取此格式。变量名限制为八字节，并自动转换成唯一的八字节名称（如果必要）。在多数情况下，不再需要以便携格式保存数据，因为 SPSS Statistics 数据文件应该独立于平台/操作系统。您无法在 Unicode 模式中以可移植文件来保存数据文件。

制表符分隔格式 (*.dat)。 用制表符分隔值的文本文件。(注意：嵌入字符串值中的 Tab 字符在制表符分隔文件中将保留为 Tab 字符。不对嵌入值中的 Tab 字符和分隔值的 Tab 字符进行区分。)

逗号分隔 (*.csv)。 用逗号或分号分隔值的文本文件。如果当前 SPSS Statistics 小数指示符为句点，则用逗号分隔各值。如果当前小数指示符为逗号，则用分号分隔各值。

固定 ASCII 格式 (*.dat)。 固定格式的文本文件，对所有变量使用缺省的书写格式。在变量字段之间没有 tab 或空格。

Excel 2007 (*.xlsx)。 Microsoft Excel 2007 XLSX 格式工作表。最大变量数为 16,000；删除超过 16,000 的任何其他变量。如果数据集包含一百万个个案，在工作表中创建多页。

Excel 97 至 2003 (*.xls)。 Microsoft Excel 97 工作表。最大变量数为 256；删除超过 256 的任何其他变量。如果数据集包含 65,356 个个案，在工作表中创建多页。

Excel 2.1 (*.xls)。 Microsoft Excel 2.1 电子表格文件。最大变量数为 256，最大行数为 16,384。

1-2-3 R3.0 (*.wk3)。 Lotus 1-2-3 电子表格文件，版本 3.0。可以保存的最大变量数为 256。

1-2-3 R2.0 (*.wk1)。 Lotus 1-2-3 电子表格文件，版本 2.0。可以保存的最大变量数为 256。

1-2-3 R1.0 (*.wks)。 Lotus 1-2-3 电子表格文件，版本 1A。可以保存的最大变量数为 256。

SYLK (*.slk)。 Microsoft Excel 和 Multiplan 电子表格文件的符号链接格式。可以保存的最大变量数为 256。

dBASE IV (*.dbf)。 dBASE IV 格式。

dBASE III (*.dbf)。 dBASE III 格式。

dBASE II (*.dbf)。 dBASE II 格式。

SAS v9+ Windows (*.sas7bdat)。 SAS v9 Windows 版。

SAS v9+ UNIX (*.sas7bdat)。 SAS v9 UNIX 版。

SAS v7-8 Windows 短扩展名 (*.sd7)。 SAS V7 – 8 for Windows 短文件名格式。

SAS v7-8 Windows 长扩展名 (*.sas7bdat)。 SAS v7 – 8 for Windows 长文件名格式。

SAS v7-8 UNIX 版 (*.sas7bdat)。 SAS v8 UNIX 版。

SAS v6 Windows 版 (*.sd2)。 用于 Windows/OS2 的 SAS V6 文件格式。

SAS v6 UNIX 版 (*.ssd01)。 用于 UNIX（Sun、HP、IBM）的 SAS V6 文件格式。

SAS v6 Alpha/OSF 版 (*.ssd04)。 用于 Alpha/OSF (DEC UNIX) 的 SAS V6 文件格式。

SAS 传输格式 (*.xpt)。 SAS 传输格式文件。

Stata V8 Intercooled (*.dta)。

Stata V8 SE (*.dta)。

Stata V7 Intercooled (*.dta)。

Stata V7 SE (*.dta)。

Stata V6 (*.dta)。

Stata V4 – 5 (*.dta)。

保存文件选项

对于电子表格、Tab 分隔文件和逗号分隔文件，您可以将变量名写入文件的第一行。

以 Excel 格式保存数据文件

您可以用三种 Microsoft Excel 文件格式之一来保存数据。Excel 2.1、Excel 97 和 Excel 2007。

- Excel 2.1 和 Excel 97 限于 256 个列；所以只能包括前 256 个变量。
- Excel 2007 限于 16,000 个列；所以只能包括前 16,000 个变量。
- Excel 2.1 限于 16,384 个行；所以只能包括前 16,384 个个案。
- Excel 97 和 Excel 2007 每页有行数的限制，但是工作表可以有多页，如果已经超过单页最大数，可以创建多页。

变量类型

下表显示 IBM® SPSS® Statistics 中的原始数据与 Excel 中的导出数据之间的变量类型匹配情况。

SPSS Statistics变量类型	Excel 数据格式
数值	0.00; #,##0.00; ...
逗号	0.00; #,##0.00; ...
美元	$#,##0_); ...
日期	d-mmm-yyyy
时间	hh:mm:ss
字符串	常规

以 SAS 格式保存数据文件

当将您的数据保存为 SAS 文件时，将对数据的各个方面进行特殊处理。这些情况包括：

- 某些在 IBM® SPSS® Statistics 变量名中允许的字符在 SAS 中是无效的，如 @、# 和 $。这些非法字符将在数据导出时用下划线来代替。
- 包含多字节字符（例如，日语或中文字符）的 SPSS Statistics 变量名将转换为一般形式的变量名，即 Vnnn，其中 nnn 是整数值。
- 包含多于 40 个字符的 SPSS Statistics 变量标签在导出至 SAS v6 文件时会被截断。
- 只要存在，SPSS Statistics 变量标签，就将它们映射为 SAS 变量标签。如果在 SPSS Statistics 数据中不存在任何变量标签，则将变量名映射为 SAS 变量标签。
- SAS 仅允许一个值为系统缺失值，而 SPSS Statistics 除系统缺失值外还允许有许多用户缺失值。因此，SPSS Statistics 中的所有用户缺失值均映射为 SAS 文件中的单个系统缺失值。

- SAS 6-8 数据文件以当前 SPSS Statistics 本地编码保存，不考虑当前模式（Unicode 或代码页）。在 Unicode 模式下，SAS 9 文件以 UTF-8 格式保存。在代码页模式下，SAS 9 文件以当前本地编码保存。
- 最多可以保存 32,767 个变量到 SAS 6-8。

保存值标签

您可以选择将与 数据文件相关联的值和值标签保存到 SAS 语法文件。该语法文件包含可在 SAS 中运行以创建 SAS 格式目录文件的 proc format 和 proc datasets 命令。

SAS 传输格式文件不支持此功能。

变量类型

下表显示 SPSS Statistics 中的原始数据与 SAS 中的导出数据之间的变量类型匹配情况。

SPSS Statistics变量类型	SAS 变量类型	SAS 数据格式
数值	数值	12
逗号	数值	12
点	数值	12
科学计数法	数值	12
日期	数值	（日期）例如：MMDDYY10 ...
日期（时间）	数值	Time18
美元	数值	12
定制货币	数值	12
字符串	字符	$8

以 Stata 格式保存数据文件

- 可以用 Stata 5 - 8 格式，也可以同时用 Intercooled 和 SE 格式（仅限版本 7 和 8）写入数据。
- 以 Stata 5 格式保存的数据文件可以由 Stata 4 读取。
- 变量标签的前 80 个字节保存为 Stata 变量标签。
- 对于数值变量，值标签的前 80 个字节保存为 Stata 值标签。对于字符串变量、非整数值和绝对值大于 2,147,483,647 的数值，将去除值标签。
- 对于版本 7 和 8，区分大小写形式的变量名的前 32 个字节保存为 Stata 变量名。对于早期版本，变量名的前 8 个字节保存为 Stata 变量名。字母、数字和下划线以外的任何字符都转换为下划线。
- 包含多字节字符（例如，日语或中文字符）的 IBM® SPSS® Statistics 变量名将转换为一般形式的变量名，即 Vnnn，其中 nnn 是整数值。

- 对于版本 5 - 6 和 Intercooled 版本 7 - 8，保存字符串值的前 80 个字节。对于 Stata SE 7 - 8，保存字符串值的前 244 个字节。
- 对于版本 5 - 6 和 Intercooled 版本 7 - 8，仅保存前 2,047 个变量。对于 Stata SE 7 - 8，仅保存前 32,767 个变量。

SPSS Statistics变量类型	Stata 变量类型	Stata 数据格式
数值	数值	g
逗号	数值	g
点	数值	g
科学计数法	数值	g
Date*、Datetime	数值	D_m_Y
Time、DTime	数值	g（秒数）
Wkday	数值	g (1 - 7)
月份	数值	g (1 - 12)
美元	数值	g
定制货币	数值	g
字符串	字符串	s

*Date、Adate、Edate、SDate、Jdate、Qyr、Moyr、Wkyr

保存变量子集

图片 3-21
“将数据保存为变量”对话框



使用“将数据保存为变量”对话框可以选择要在新数据文件中保存的变量。缺省情况下将保存所有变量。取消选择您不想保存的变量，或单击**全部丢弃**然后选择想要保存的变量。

仅可视。仅选择当前使用的变量集中的变量。

保存变量子集

▶ 使数据编辑器成为活动窗口（单击窗口的任何部位即可使其成为活动窗口）。

▶ 从菜单中选择：
文件 > 另存为...

▶ 单击**变量**。

▶ 选择想要保存的变量。

导出到数据库

可使用导出到数据库向导完成以下任务：

- 替换现有数据库表字段（列）中的值或为表添加新字段。
- 将新记录（行）追加到数据库表。
- 完全替换数据库表或创建新表。

要将数据导出到数据库，请执行下列操作：

▶ 在包含要导出数据的数据集的“数据编辑器”窗口中，从菜单中选择：
文件 > 导出到数据库

▶ 选择数据库源。

▶ 按照导出向导中的说明操作以导出数据。

从 IBM SPSS Statistics 变量创建数据库字段

创建新字段（向现有数据库表添加字段、创建新表或替换表）时，可以指定字段名、数据类型和宽度（适用的情况下）。

字段名称。缺省字段名称与 IBM® SPSS® Statistics 变量名相同。可以将字段名称更改为数据库格式允许的任何名称。例如，很多数据库允许字段名称中包含 变量名中不允许的字符，包括空格。因此，类似于 CallWaiting 的 变量名可以更改为字段名称 Call Waiting。

类型。导出向导基于标准 ODBC 数据类型或选定的数据库格式允许的与定义的 SPSS Statistics 数据格式最匹配的数据类型进行首次数据类型指定，但是数据库可对在 SPSS Statistics 中没有直接对应类型的类型进行区分。例如，SPSS Statistics 中的很多数值都以双精度浮点值保存，而数据库数值类型包含浮点（双精度）、整数、实数等等。此外，很多数据库没有与 SPSS Statistics 对应的时间格式类型。可以将数据类型更改为下拉列表中可用的任何类型。

作为一个通常的规则，变量的基本数据类型（字符串或数值）应与数据库字段的基本数据类型相匹配。如果出现了数据库无法解决的数据类型不匹配，则会出现错误结果，并且不会将任何数据导出到数据库中。例如，如果将 字符串变量导出到数值数据类型的数据库字段，则当字符串变量的任何值包含非数字字符时，结果将出错。

宽度。可以对 string (char, varchar) 字段类型的已定义宽度进行更改。数字字段宽度由数据类型决定。

缺省情况下，SPSS Statistics 变量格式根据下列总体原则映射到数据库字段类型。实际数据库字段类型可能取决于数据库。

SPSS Statistics 变量格式	数据库字段类型
数值	浮点数或双精度数
逗号	浮点数或双精度数
点	浮点数或双精度数
科学计数法	浮点数或双精度数
日期	日期、日期时间或时间戳
日期时间	日期时间或时间戳
Time、DTime	浮点数或双精度数（秒数）
Wkday	整数 (1 - 7)
月份	整数 (1 - 12)
美元	浮点数或双精度数
定制货币	浮点数或双精度数
字符串	Char 或 Varchar

用户缺失值

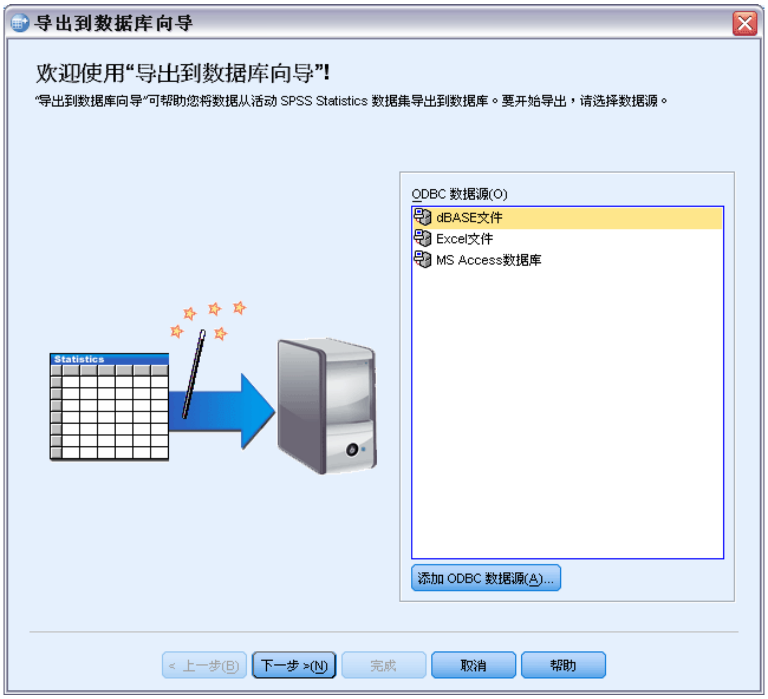
将来自变量的数据导出到数据库字段时，有两个选项可用于处理用户缺失值：

- **导出为有效值**。用户缺失值当作常规的、有效的非缺失值处理。
- **将数值型的用户缺失值作为 Null 导出，并将字符串类型的用户缺失值作为空格导出**。数值型的用户缺失值被视为与系统缺失值相同。字符串类型的用户缺失值被转换为空格（字符串不能是系统缺失的）。

选择数据源

在导出到数据库向导的第一个面板中，选择要向其导出数据的数据源。

图片 3-22
导出到数据库向导，选择数据源



可以将数据导出到具有相应 ODBC 驱动程序的任何数据库源。(注意：不支持将数据导出到 OLE DB 数据源。)

如果没有配置任何 ODBC 数据源，或者要添加新的数据源，请单击**添加 ODBC 数据源**。

- 在 Linux 操作系统中，该按钮不可用。在 odbc.ini 中指定 ODBC 数据源，并且 ODBCINI 环境变量必须设定为该文件的位置。有关更多信息，请参见数据库驱动程序文档。
- 用分布式分析模式时（IBM® SPSS® Statistics Server 提供），该按钮不可用。要用分布式分析模式添加数据源，请咨询系统管理员。

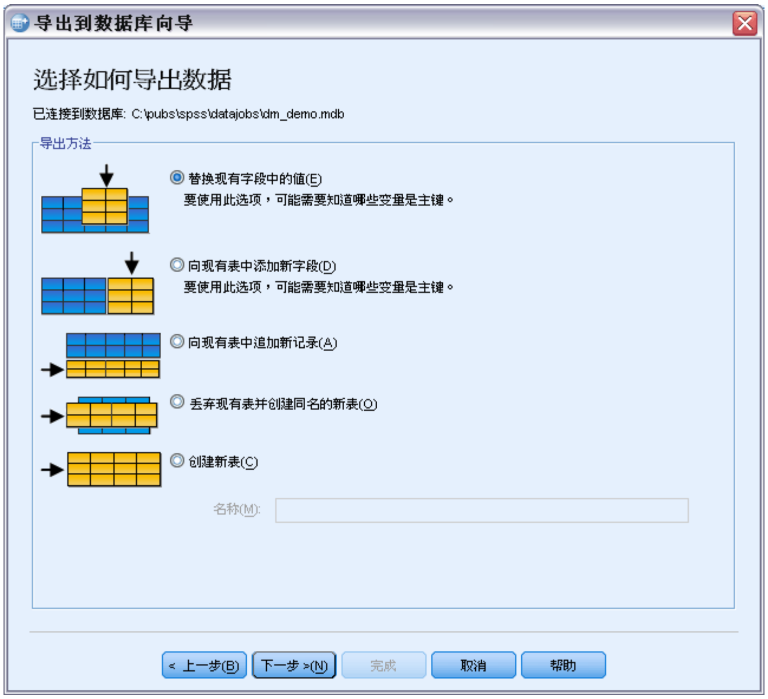
ODBC 数据源由两条基本信息组成：要用来访问数据的驱动程序和要访问的数据库的位置。要指定数据源，必须装有适当的驱动程序。针对不同数据库格式的驱动程序可以从 http://www.spss.com/drivers 获得。

有些数据源可能要求登录 ID 和密码才能进行到下一步。

选择如何导出数据

选择了数据源后，需要指明导出数据所采用的方式。

图片 3-23
导出到数据库向导，选择导出方式



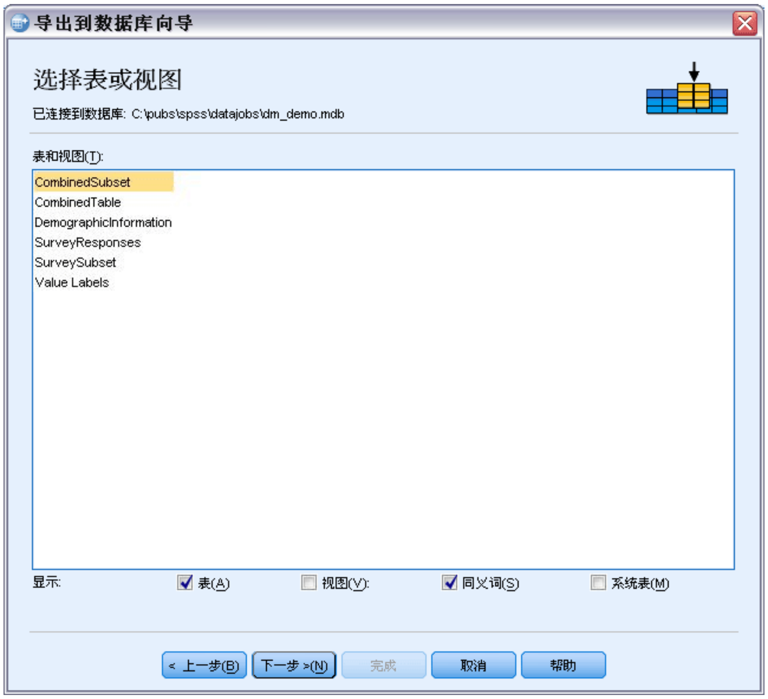
以下选择可供将数据导出到数据库：

- **替换现有字段中的值。**将现有表中选定字段的值用活动数据集中选定变量的值替换。
- **向现有表中添加新字段。**在现有表中创建新字段，这些字段将包含活动数据集中选定变量的值。此选项不支持 Excel 文件。
- **向现有表中追加新记录。**向现有表添加新记录（行），这些记录将包含活动数据集中的个案的值。
- **丢弃现有表并创建同名的新表。**删除指定的表并创建同名的新表，新表中将包含活动数据集中的选定变量。原始表中包括字段属性定义（例如主键和数据类型）在内的所有信息都将丢失。
- **创建新表。**在数据库中创建新表，其中包含来自活动数据集中的选定变量的数据。名称可以是数据源允许作为表名的任何值。此名称不能与数据库中现有表或视图名称重复。

选择表

修改或替换数据库中的表时，需要选择要修改或替换的表。导出到数据库向导中的此面板显示了一个列表，其中列出选定数据库中的表和视图。

图片 3-24
导出到数据库向导，选择表或视图



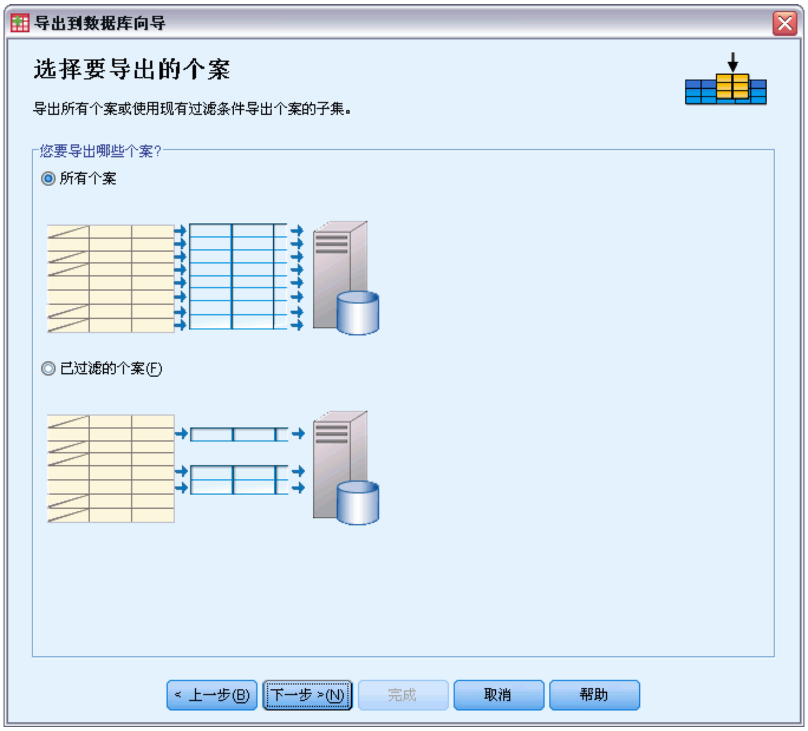
默认情况下，此列表仅显示标准数据库表。可以控制列表中显示的项的类型：

- **表**。标准数据库表。
- **视图**。视图是由查询定义的虚拟的或动态的“表”。视图中可以包含基于其他字段值计算得出的多个表和/或字段的连接。虽然可以对视图追加记录，或替换其现有字段的值，但是您能够修改的字段是受限制的，这取决于视图的结构。例如，不能修改派生的字段、为视图添加字段或替换视图。
- **同义词**。同义词是表或视图的别名，通常在查询中定义。
- **系统表**。系统表定义数据库的属性。在某些情况下，标准数据库表可能会被归类为系统表，并且仅在选择了该选项后才会显示。通常只有数据库管理员才有访问真正的系统表的权限。

选择要导出的个案

导出到数据库向导中对个案的选择仅限于所有个案以及使用先前定义的过滤条件选择的个案。如果没有有效的个案过滤，此面板将不显示，并将导出活动数据集中的所有个案。

图片 3-25
导出到数据库向导，选择要导出的个案



有关定义用于个案选择的过滤器条件的信息，请参见选择个案第 169 页码。

将个案匹配到记录

在向现有表添加字段（列）或替换现有字段中的值时，需确保活动数据集中的每个个案（行）与数据库中相应的记录能够正确匹配。

- 在数据库中，唯一标识每个记录的字段或字段组通常被指定为**主键**。
- 需要确定与主键字段或其他能唯一标识每个记录的字段组对应的变量。
- 字段不一定必须是数据库中的主键，但是字段值或字段值的组合对于每个个案必须是唯一的。

要使变量与数据库中唯一标识每个记录的字段匹配，请执行下列操作：

▶ 将变量拖放至相应的数据库字段上。

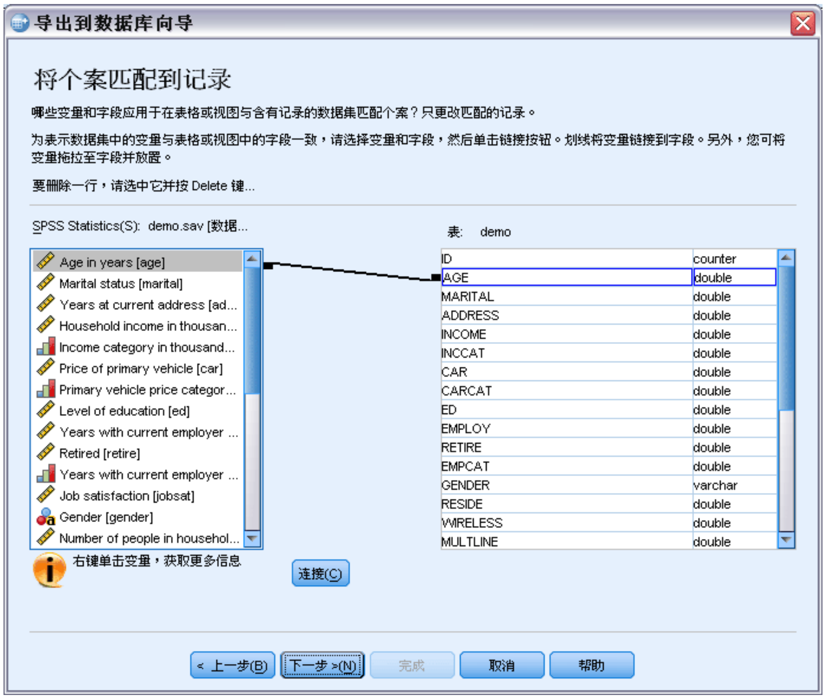
或

▶ 从变量列表中选择变量，再选择数据库表中相应的字段，然后单击**连接**。

要删除连接线，请执行下列操作：

▶ 选择连接线并按 Delete 键。

图片 3-26
导出到数据库向导，将个案匹配到记录



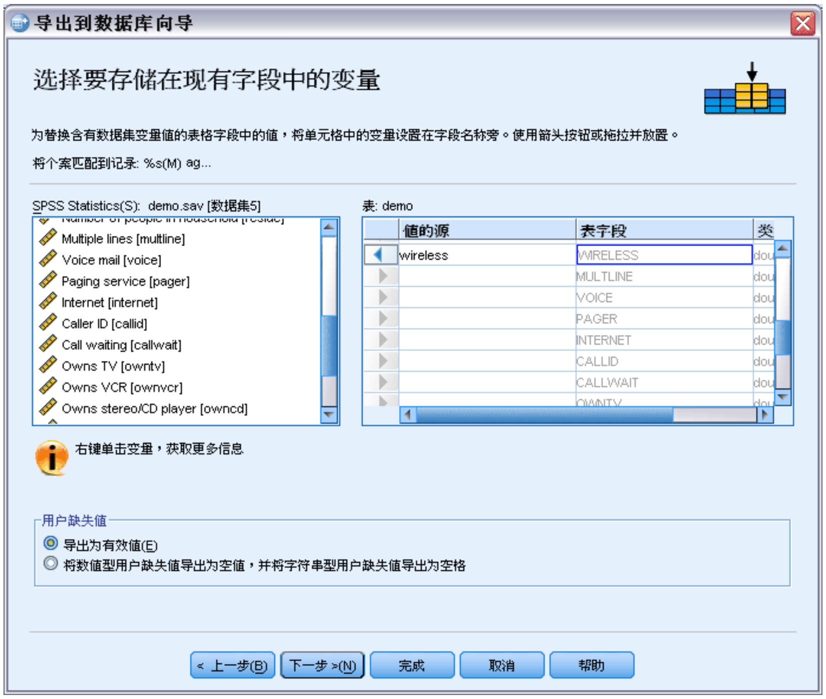
注意：变量名和数据库字段名不一定相同（因为数据库字段名可能包含 IBM® SPSS® Statistics 变量名所不允许的字符），但是如果活动数据集是从所修改的数据库表创建的，那么变量名或变量标签两者之一通常将至少与数据库字段名相似。

替换现有字段中的值

要替换数据库中现有字段的值，请执行下列操作：

- 在导出到数据库向导的**选择如何导出数据**面板中，选择**替换现有字段中的值**。
- 在**选择表或视图**面板中，选择数据库表。
- 在**将个案匹配到记录**面板中，将唯一标识每个个案的变量与相应的数据库字段名称相匹配。
- 对于要替换其值的每个字段，将包含新值的变量拖放到相应数据库字段名旁的**值的源**列。

图片 3-27
导出到数据库向导，替换现有字段中的值



- 作为一个通常的规则，变量的基本数据类型（字符串或数值）应与数据库字段的基本数据类型相匹配。如果出现了数据库无法解决的数据类型不匹配，则会出现错误结果，并且不会将任何数据导出到数据库中。例如，如果将字符串变量导出到数值数据类型（例如双精度数、实数或整数）的数据库字段，则当字符串变量的任何值包含非数字字符时，结果将出错。变量旁的图标中的字母 a 指示该变量为字符串变量。
- 不能修改字段名、类型或宽度。原始数据库字段属性将被保留，只有值被替换。

添加新字段

要向现有数据库表添加新字段，请执行下列操作：

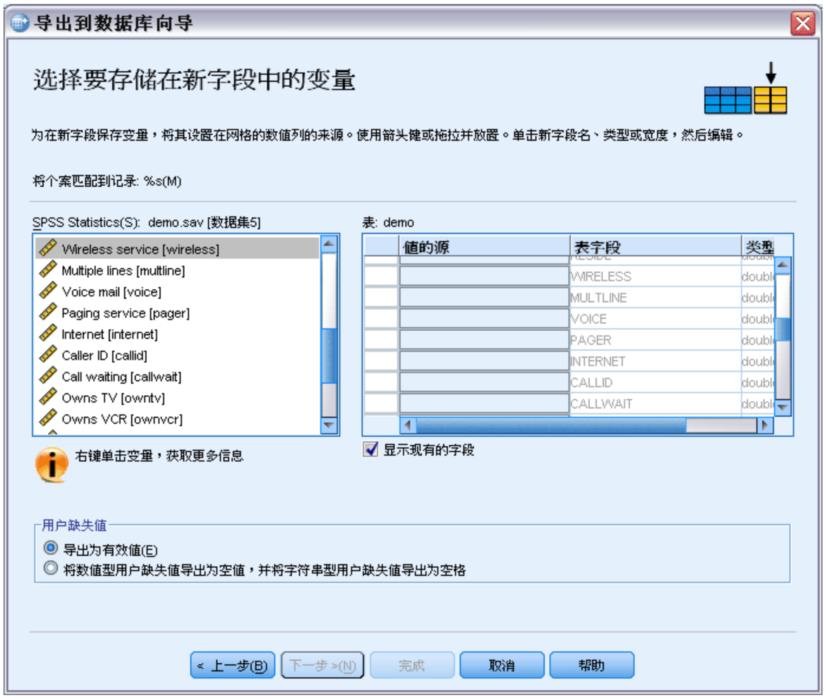
▶ 在导出到数据库向导的**选择如何导出数据**面板中，选择**向现有表中添加新字段**。

▶ 在**选择表或视图**面板中，选择数据库表。

▶ 在**将个案匹配到记录**面板中，将唯一标识每个个案的变量与相应的数据库字段名称相匹配。

▶ 将要作为新字段添加的变量拖放到**值的源**列。

图片 3-28
导出到数据库向导，向现有表中添加新字段



有关字段名称和数据类型的信息，请参见导出到数据库第 44 页码中有关从 IBM® SPSS® Statistics 变量创建数据库字段的章节。

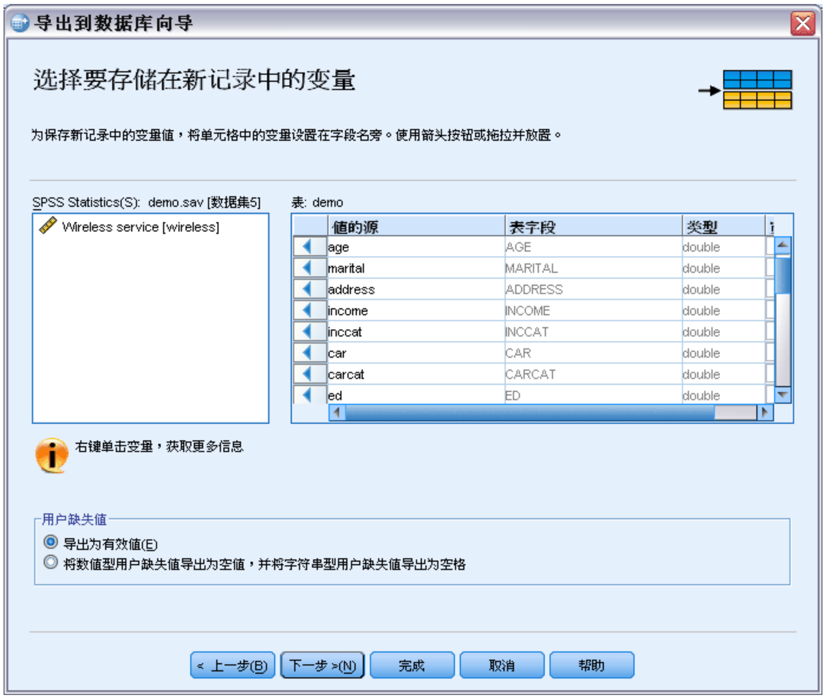
显示现有的字段。选择此选项以显示现有字段的列表。不能使用导出到数据库向导中的此面板替换现有字段，但是了解表中已有的字段会有帮助。如果要替换现有字段中的值，请参见替换现有字段中的值第 50 页码。

追加新记录（个案）

要将新记录（个案）追加到数据库表，请执行下列操作：

- 在导出到数据库向导的**选择如何导出数据**面板中，选择**向现有表中追加新记录**。
- 在**选择表或视图**面板中，选择数据库表。
- 将活动数据集中的变量与表字段匹配，方法是将变量拖放到**值的源**列。

图片 3-29
导出到数据库向导，向表添加记录（个案）



导出到数据库向导将根据存储在活动数据集（如可用）中的原始数据库表和/或与字段名同名的变量名的信息，自动选择与现有字段匹配的所有变量。此初始自动匹配功能的目的仅是提供一个指导，并不妨碍您更改变量与数据库字段匹配的方式。

向现有表中添加新字段时，将应用以下基本规则/限制：

- 活动数据集中的所有个案（或所有选定个案）都将添加到表中。如果这些个案中任何一个与数据库中的现有记录重复，则当遇到重复键值时，会发生错误。有关仅导出选定个案的信息，请参见选择要导出的个案第 48 页码。
- 可以使用会话中创建的新变量的值作为现有字段的值，但是不能添加新字段或更改现有字段的名称。要向数据库表添加新字段，请参见添加新字段第 51 页码。
- 对于任何排除的数据库字段或未匹配到变量的字段，数据库表中的新增记录将没有相应的值。（如果**值的源**单元格为空，则没有任何变量匹配到字段。）

创建新表或替换表

要创建新的数据库表，或替换现有的数据库表，请执行下列步骤：

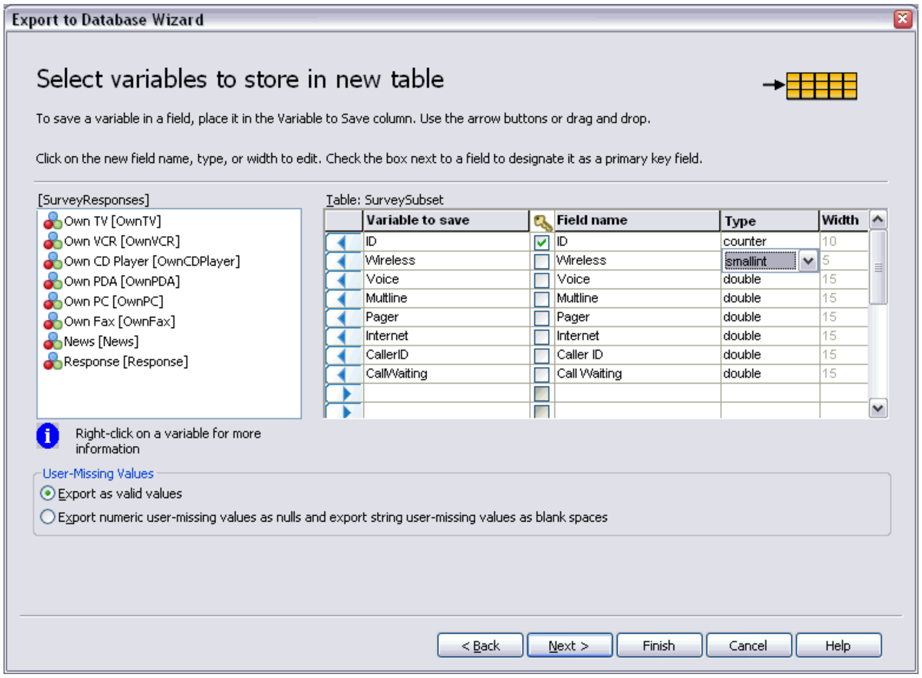
▶ 在导出向导的**选择如何导出数据**面板中，选择**丢弃现有表并创建同名的新表**或选择**创建新表**并为新表输入一个名称。如果表名称包含除字母、数字或下划线以外的任何其他字符，名称必须用双引号括起。

▶ 如果要替换现有的表，则在**选择表或视图**面板中选择数据库表。

▶ 将变量拖放到**要保存的变量**列。

▶ （可选）您可以指定定义主键的变量/字段，更改字段名以及更改数据类型。

图片 3-30
导出到数据库向导，为新表选择变量



主键。要将变量指定为数据库表中的主键，请选择由键图标标识的列中的框。

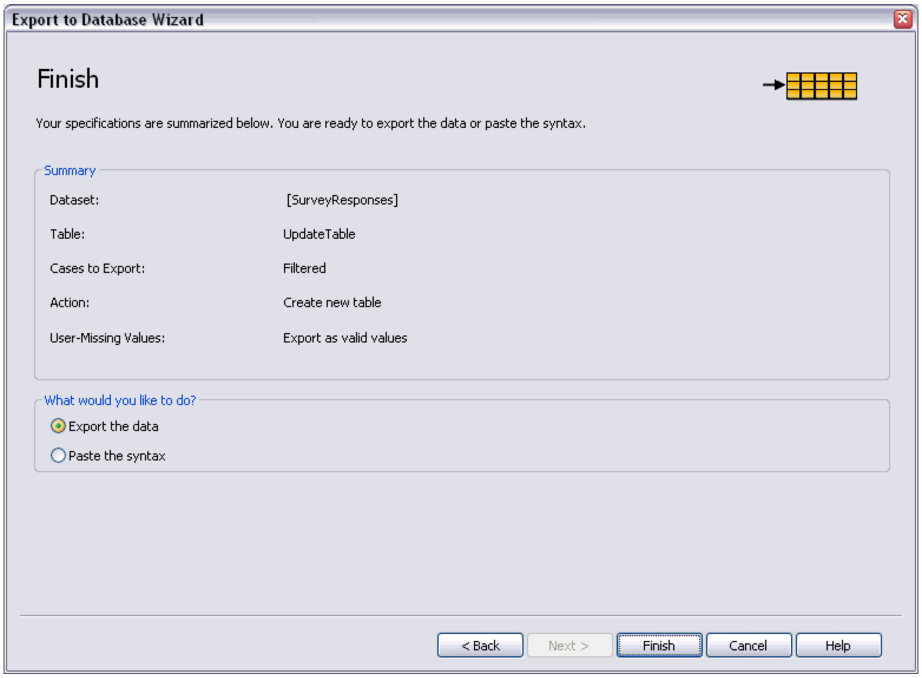
- 主键的所有值必须是唯一的，否则将发生错误。
- 如果选择了单个变量作为主键，则每个记录（个案）对于该变量必须具有唯一值。
- 如果选择了多个变量作为主键，这定义了组合主键，则选定变量的值的组合对每个个案来说必须是唯一的。

有关字段名称和数据类型的信息，请参见导出到数据库第 44 页码中有关从 IBM® SPSS® Statistics 变量创建数据库字段的章节。

完成数据库导出向导

“数据库导出向导”的最后一个面板提供了摘要，其中指明要导出的数据和导出方式。此面板还提供了是将数据导出还是将底层命令语法粘贴到语法窗口的选项。

图片 3-31
导出到数据库向导，“完成”面板



摘要信息

- **数据集。**将用于导出数据的数据集的 IBM® SPSS® Statistics 会话名。如果有多个打开的数据源，此信息将非常有用。对于通过图形用户界面（如数据库向导）打开的数据源，将自动指定诸如 DataSet1、DataSet2 等的名称。通过命令语法打开的数据源将只有数据集名（如果明确指定了一个名称）。
- **表**要修改或创建的表的名称。
- **要导出的个案。**导出所有个案，或导出由先前定义的过滤器条件选定的所有个案。
- **操作。**表示对数据库的修改方式（例如，创建新表或者向现有表添加字段或记录）。
- **用户缺失值。**用户缺失值可作为有效值导出；或者，对于数值变量与系统缺失值同等对待，对于字符串变量转换为空格。此设置在选择要导出的变量的面板中控制。

导出到 IBM SPSS Data Collection

“导出到 IBM® SPSS® Data Collection”对话框可以创建 IBM® SPSS® Statistics 数据文件和 Data Collection 元数据文件，您可以使用这些文件将数据读取到 Data Collection 应用程序中。在 SPSS Statistics 和 Data Collection 应用程序之间“穿插”使用数据时，该操作非常有用。

要将数据导出以供 Data Collection 应用程序使用，请执行下列操作：

▶ 在包含要导出的数据的“数据编辑器”窗口中，从菜单中选择：

文件 > 导出到 Data Collection

▶ 单击**数据文件**以指定 SPSS Statistics 数据文件的名称和位置。

▶ 单击**元数据文件**以指定 Data Collection 元数据文件的名称和位置。

对于不是从 Data Collection 数据源创建的新变量和数据集，则根据 IBM® SPSS® Data Collection Developer Library 中的 SAV DSC 文档中描述的方法，SPSS Statistics 变量特性将映射到元数据文件中的 Data Collection 元数据特性。

如果活动数据集是从 Data Collection 数据源创建的：

- 新的元数据文件是通过将原始元数据库特性与任何新变量的元数据特性合并而来的，再加上可能影响其元数据特性（例如添加或更改值标签）的对原始变量的任何更改。
- 对于从 Data Collection 数据源读取的原始变量，SPSS Statistics 不识别的任何元数据特性都将保留其原始状态。例如，SPSS Statistics 将网格变量转换为常规 SPSS Statistics 变量，但定义这些网格变量的元数据将在您保存新元数据文件时保留。
- 如果任何 Data Collection 变量自动重命名以符合 SPSS Statistics 变量命名规则，则元数据文件会将已转换的名称映射到原始 Data Collection 变量名。

值标签是否存在会影响变量的元数据特性，并因此影响 Data Collection 应用程序读取这些变量的方式。如果已为变量的任何非缺失值定义了值标签，则应为该变量的所有非缺失值定义值标签；否则，当 Data Collection 读取数据文件时，未加标签的值将丢失。

此功能仅在安装了 SPSS Statistics 的 Microsoft Windows 操作系统上可用，且仅在本地分析模式中可用。此功能不能用于使用 SPSS Statistics Server 的分布式分析模式。

要写入 Data Collection 元数据文件，必须安装以下项目：

- .NET Framework。要获得最新版本的 .NET Framework，请转至 http://www.microsoft.com/net。
- IBM® SPSS® Data Collection Survey Reporter Developer Kit。有关获取 SPSS Survey Reporter Developer Kit 兼容版本的信息，请转至 support.spss.com (http://support.spss.com)。

保护原始数据

为了防止原始数据被意外修改或删除，您可以将该文件标记为只读。

▶ 从“数据编辑器”菜单中选择：
文件 > 将文件标记为只读

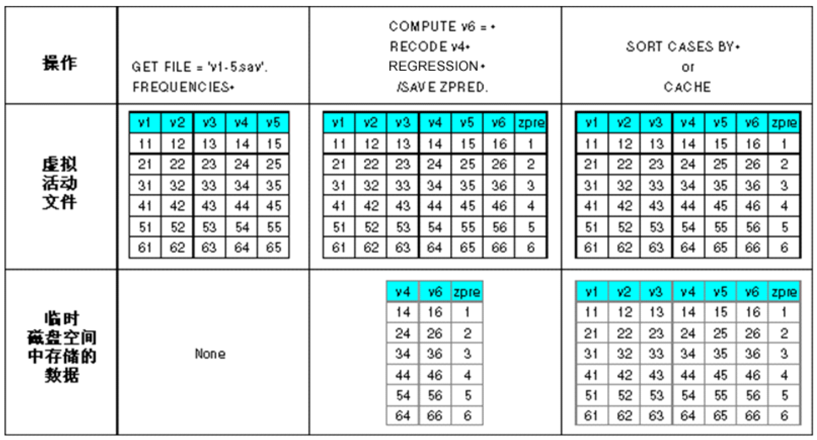
如果随后对数据进行修改，然后尝试保存数据文件，则只能用其他文件名保存数据；这样原始数据就不会受影响。

通过从“文件”菜单中选择**将文件标记为读写**您可以将文件权限改回为读写。

虚拟活动文件

虚拟活动文件使您能够处理大型数据文件，而无需足够大（或更大）的临时磁盘空间。对于大多数分析和绘图过程，会在每次您运行不同的过程时重新读取初始数据源。修改数据的过程需要确定的临时磁盘空间量来记录更改，而且一些操作始终需要足够的磁盘空间容纳数据文件的至少一个完整副本。

图片 3-32
临时磁盘空间要求



不需要任何临时磁盘空间的操作包括：

- 读取 IBM® SPSS® Statistics 数据文件
- 合并两个或更多 SPSS Statistics 数据文件
- 使用数据库向导读取数据库表
- 将 SPSS Statistics 数据文件与数据库表合并
- 运行读取数据的过程（例如“频率”、“交叉表”和“探索”）

在临时磁盘空间中创建一列或多列数据的操作包括：

- 计算新变量
- 对现有变量重新编码
- 运行创建或修改变量的过程（例如在“线性回归”中保存预测值）

在临时磁盘空间中创建数据文件的整个副本的操作包括：

- 读取 Excel 文件
- 运行对数据排序的过程（例如“对个案进行排序”和“拆分文件”）
- 使用 GET TRANSLATE 或 DATA LIST 命令读取数据
- 使用“高速缓存数据”工具或 CACHE 命令
- 从 SPSS Statistics 启动其他读取数据文件的应用程序（例如 AnswerTree 和 DecisionTime）

注意：GET DATA 命令提供类似于 DATA LIST 的功能，但不在临时磁盘空间中创建数据文件的完整副本。命令语法中的 SPLIT FILE 命令不对数据文件进行排序，因此不创建数据文件的副本。但是此命令需要已排序的数据才能正确运行，此过程的对话框接口将自动对数据文件进行排序，生成数据文件的完整副本。（命令语法在学生版中不可用。）

缺省情况下创建数据文件的完整副本的操作：

- 使用数据库向导读取数据库
- 使用文本向导读取文本文件

文本向导提供了可选的设置以自动高速缓存数据。缺省情况下，此选项是选中的。您可以通过取消选择**在本地高速缓存数据**关闭此选项。对于数据库向导，您可以粘贴生成的命令语法并删除 CACHE 命令。

创建数据高速缓存

尽管虚拟活动文件可以显著地减少所需的临时磁盘空间量，但是缺少“活动”文件的临时副本意味着必须为每个过程均重复读取初始数据源。对于从外部源读取的大型数据文件，创建数据的临时副本可以改善性能。例如，对于从数据库源读取的数据表，必须为需要读取数据的任何命令或过程重复执行从数据库读取信息的 SQL 查询。由于实际上所有统计分析过程和图表绘制过程均需要读取数据，因此会为您运行的每个过程重复执行 SQL 查询，如果您运行大量过程，这会导致处理时间的显著增加。

如果您在执行分析的计算机（本地计算机或远程计算机）上有足够的磁盘空间，则可以通过创建活动文件的数据高速缓存来消除多次 SQL 查询，并减少处理时间。此数据高速缓存是完整数据的临时副本。

注意：缺省情况下，数据库向导自动创建数据高速缓存，但是如果您在命令语法中使用 GET DATA 命令读取数据库，则不会自动创建数据高速缓存。（命令语法在学生版中不可用。）

创建数据高速缓存

▶ 从菜单中选择：
文件 > 高速缓存数据...

▶ 单击**确定**或**立即高速缓存**。

单击**确定**会在程序下次读取数据（例如，下次您运行统计程序）时创建数据高速缓存，这通常是您想要的，因为它不要求额外的数据传输。**立即高速缓存**会立即创建数据高速缓存，这在大多数情况下是不必要的。**立即高速缓存**主要用于以下两种情况：

- 数据源被“锁定”，在您结束会话、打开不同的数据源或高速缓存数据之前，任何人都无法更新该数据源。
- 对于大型数据源，如果您高速缓存了数据，则在数据编辑器的“数据视图”选项卡的内容中进行滚动会快得多。

自动高速缓存数据

可以使用 SET 命令，在活动数据文件中发生指定数量的更改之后自动创建数据高速缓存。缺省情况下，在活动数据文件发生 20 个更改之后，会自动高速缓存该活动数据文件。

▶ 从菜单中选择：
文件 > 新建 > 语句

▶ 在语法窗口中键入 SET CACHE n（其中 n 表示数据文件高速缓冲前活动数据文件中更改的数量）。

▶ 从语法窗口中的菜单选择：
运行 > 全部

注意：高速缓存设置并不是在各个会话之间持续存在。每次启动新会话时，该值均重置为缺省值 20。

章

4

分布式分析模式

分布式分析模式允许您使用本地（或桌面）计算机以外的计算机以进行内存密集型工作。由于用于分布式分析的远程服务器通常比本地计算机性能更强，速度更快，因此使用分布式分析模式可以显著地缩短计算机处理时间。如果工作涉及到以下内容，则通过远程服务器进行分布式分析可能会有用：

- 大型数据文件，尤其是从数据库源中读取的数据。
- 内存密集型任务。任何在本地分析模式下耗时很长的任务均是适合于进行分布式分析的对象。

分布式分析仅影响与数据相关的任务，例如读取数据、转换数据、计算新变量和计算统计量。分布式分析对与编辑输出相关的任务没有影响，例如操作枢轴表或修改图表。

注意：只有在既有该软件的本地版本，又能访问安装在远程服务器上的该软件的许可服务器版本的情况下，才能使用分布式分析。

服务器登录

“服务器登录”对话框允许您选择处理命令和运行过程的计算机。您可以选择本地计算机，也可以选择远程服务器。

图片 4-1
“服务器登录”对话框



您可以在列表中添加、修改或删除远程服务器。远程服务器通常需要用户标识和密码，可能还需要域名。请联系系统管理员以获取关于可用服务器、用户标识和密码、域名的信息以及其他连接信息。

您可以选择缺省服务器并保存与任何服务器相关联的用户标识、域名和密码。启动新会话时，会自动连接到缺省服务器。

如果您被许可使用 Statistics Adapter，且站点正在运行 IBM® SPSS® Collaboration and Deployment Services 3.5 或更新版本，则可单击**搜索...**以查看网络上可用的服务器列表。如果尚未登录到 IBM® SPSS® Collaboration and Deployment Services Repository，则您将被提示输入连接信息，然后才能查看服务器列表。

添加或编辑服务器登录设置

使用“服务器登录设置”对话框可以添加或编辑远程服务器的连接信息，以便在分布式分析模式下使用。

图片 4-2
“服务器登录设置”对话框



请联系系统管理员获取可用服务器列表、服务器端口号以及其他连接信息。除非有管理员指示，否则请勿使用“安全套接字层”。

服务器名称。服务器?名称?可以是分配给计算机的字母数字名称（例如 NetworkServer），也可以是分配给计算机的 IP 地址（例如 202.123.456.78）。

端口号。端口号是服务器软件用于通信的端口。

描述。可以输入可选的描述以显示在服务器列表中。

使用安全套接字层连接。安全套接字层 (SSL) 在分布式分析请求发送到远程服务器时加密请求。使用 SSL 之前，请与管理员协商。要启用此选项，必须在桌面计算机和服务器上配置 SSL。

选择、切换或添加服务器

- 从菜单中选择：
  文件 > 切换服务器...

选择缺省服务器：

- 在服务器列表中，选择要使用的服务器旁边的框。
- 输入管理员提供的用户标识、域名和密码。

  注意：启动新会话时，会自动连接到缺省服务器。

切换到其他服务器：

- 从列表中选择服务器。
- 输入用户标识、域名和密码（如果需要）。

  注意：在会话期间切换服务器时，所有打开的窗口均会关闭。在窗口关闭之前会提示您保存更改。

添加服务器：

- 从管理员处获取服务器连接信息。
- 单击**添加**打开?服务器登录设置?对话框。
- 输入连接信息和可选设置，然后单击**确定**。

编辑服务器：

- 从管理员处获取修订的连接信息。
- 单击**编辑**打开?服务器登录设置?对话框。
- 输入更改并单击**确定**。

选择缺省服务器：

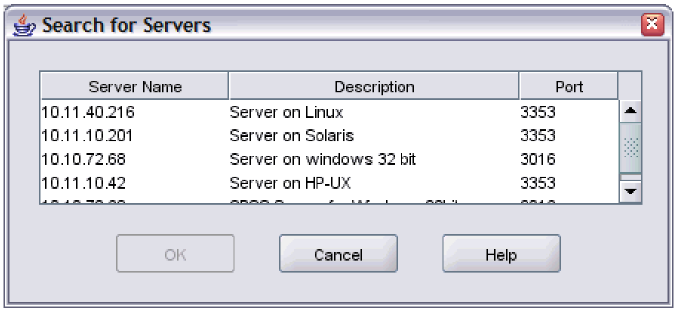
注意：仅当您被许可使用 Statistics Adapter，且站点正在运行 IBM® SPSS® Collaboration and Deployment Services 3.5 或更新版本时，才可使用搜索可用服务器功能。

- 单击**搜索...**以打开?搜索服务器?对话框。如果您尚未登录到 IBM® SPSS® Collaboration and Deployment Services Repository，将提示您连接信息。
- 选择一个或多个服务器，并单击**确定**。服务器现在将出现在“服务器登录”对话框中。
- 要连接到其中一个服务器，请按照“切换到其他服务器”的说明进行操作。

搜索可用服务器

使用?搜索服务器?对话框，选择网络上可用的一个或多个服务器。在您从?服务器登录?对话框中单击**搜索...**时，该对话框出现。

图片 4-3
“搜索服务器”对话框



选择一个或多个服务器，单击**确定**以将其添加到?服务器登录?对话框中。尽管您可在“服务器登录”对话框中手动添加服务器，但搜索可用服务器功能允许您无需知道正确的服务器名称和端口号，就能连接到服务器。这信息会自动提供。不过，您仍然需要提供正确的登录信息，例如用户名、域名和密码等。

从远程服务器打开数据文件

在分布式分析模式下，“打开远程文件”对话框替换了标准的“打开文件”对话框。

- 可用文件、文件夹和驱动器的列表内容取决于远程服务器上可用的，或可从该服务器访问的文件、文件夹和驱动器。当前服务器名称在对话框的顶部指明。
- 在分布式分析模式下，除非将驱动器指定为共享设备，或者将包含数据文件的文件夹指定为共享文件夹，否则将不能访问本地计算机上的数据文件。有关如何与服务器网络“共享”本地计算机上的文件夹的信息，请参见操作系统文档。
- 如果服务器在运行不同的操作系统（例如您在运行 Windows 而服务器在运行 UNIX），那么即使本地数据文件位于共享文件夹中，您在分布式分析模式下也可能无法访问它们。

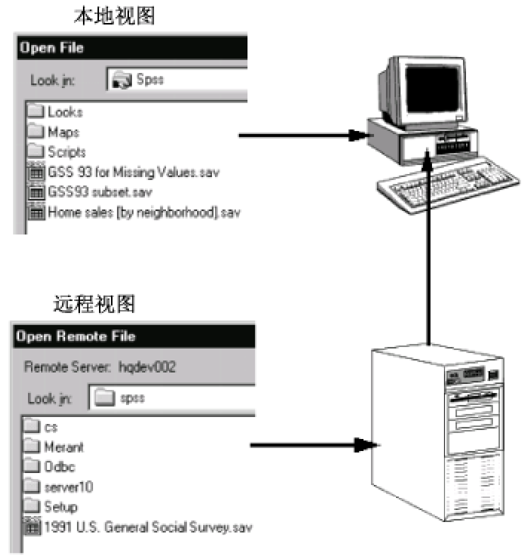
本地和分布式分析模式下的文件访问

本地计算机和网络的数据文件、文件夹（目录）和驱动器视图基于您正用来处理命令和运行过程的计算机，该计算机不一定是您面前的计算机。

本地分析模式。如果使用本地计算机作为您的?服务器?，则您在文件访问对话框中看到的数据文件、文件夹和驱动器视图（对于打开数据文件）类似于在其他应用程序或在 Windows 资源管理器中看到的视图。您可以看到您的计算机上的所有数据文件和文件夹，以及所安装的网络驱动器上的所有文件和文件夹。

分布式分析模式。如果使用另外一台计算机作为运行命令和过程的?远程服务器?，则数据文件、文件夹和驱动器的视图代表从远程服务器计算机看到的视图。尽管您可以看到熟悉的文件夹名称（例如 Program Files）和驱动器名称（例如 C），但这些不是 您的计算机上的文件夹或驱动器；它们是远程服务器上的文件夹和驱动器。

图片 4-4
本地和远程视图



在分布式分析模式下，除非将驱动器指定为共享设备，或者将包含数据文件的文件夹指定为共享文件夹，否则将不能访问本地计算机上的数据文件。如果服务器在运行不同的操作系统（例如您在运行 Windows 而服务器在运行 UNIX），那么即使本地数据文件位于共享文件夹中，您在分布式分析模式下也可能无法访问它们。

分布式分析模式不同于访问驻留在网络上另一台计算机上的数据文件。在本地分析模式或分布式分析模式下您都可以访问其他网络设备上的数据文件。在本地模式下，您是从本地计算机访问其他设备。而在分布式模式下，您是从远程服务器访问其他网络设备。

如果对于是在使用本地分析模式还是分布式分析模式没有把握，可查看用于访问数据文件的对话框中的标题栏。如果对话框的标题包含词远程（例如**打开远程文件**），或者如果文本**远程服务器**:[**服务器名称**] 显示在对话框顶部，则说明您在使用分布式分析模式。

注意：此情况只对访问数据文件的对话框（例如“打开数据”、“保存数据”、“打开数据库”和“应用数据字典”）有影响。对于所有其他文件类型（例如查看器文件、语法文件和脚本文件），始终使用本地视图。

分布式分析模式下过程的可用性

在分布式分析模式下，只有同时安装在本地版本和远程服务器上的版本中的过程才可用。

如果在本地安装远程服务器上未安装的可选组件，并从本地计算机切换到远程服务器，则受影响的程序将从菜单中移去，相应的命令语法将导致出现错误。切换回本地方式可恢复所有受影响的过程。

绝对和相对路径指定

在分布式分析模式下，数据文件和命令语法文件的相对路径指定是相对于当前服务器的，而不是相对于本地计算机。相对路径指定（如 /mydocs/mydata.sav）并不指向您的本地驱动器上的目录和文件；它指向远程服务器的硬盘上的目录和文件。

Windows UNC 路径指定

如果在使用 Windows 服务器版本的程序，则在使用命令语法访问数据和语法文件时，可以使用通用命名约定 (UNC) 指定。UNC 指定的一般形式为：

```
\\servername\sharename\path\filename
```

- Servername 是包含数据文件的计算机的名称。
- Sharename 是该计算机上指定为共享文件夹的文件夹（目录）。
- Path 是共享文件夹下的任何其他文件夹（子目录）路径。
- Filename 是数据文件的名称。

示例如下：

```
GET FILE='\\hqdev001\public\july\sales.sav'。
```

如果计算机没有为其指定的名称，则可以使用其 IP 地址，例如：

```
GET FILE='\\204.125.125.53\public\july\sales.sav'。
```

即使是通过 UNC 路径指定，您也只能访问指定为共享的驱动器和文件夹中的数据与语法文件。使用分布式分析模式时，此情况包括本地计算机上的数据与语法文件。

UNIX 绝对路径指定

在 UNIX 服务器版本中，不存在与 UNC 路径等同的路径指定法，所有目录路径必须是以服务器根目录开始的绝对路径，不允许使用相对路径。例如，如果数据文件位于 /bin/data 且当前目录也是 /bin/data，则GET FILE='sales.sav' 无效；必须指定整个路径，例如：

```
GET FILE='/bin/sales.sav'.
INSERT FILE='/bin/salesjob.sps'.
```

章
5

数据编辑器

数据编辑器提供一种类似电子表格的便利方法来创建和编辑数据文件。当您启动会话时，“数据编辑器”窗口自动打开。

数据编辑器提供数据的两种视图：

- **数据视图。**显示实际的数据值或定义的值标签。
- **变量视图。**显示变量定义信息，包括定义的变量标签和值标签、数据类型（例如，字符串、日期或数值）、测量级别（名义、序数或刻度）及用户定义的缺失值。

在两种视图中，您都可以添加、更改和删除包含在数据文件中的信息。

数据视图

图片 5-1
数据视图



数据视图的许多功能类似于电子表格应用程序中提供的功能。不过，有一些重要的不同之处：

- 行为个案。每一行代表一个个案或一个观察值。例如，问卷的每个回答者都是一个个案。

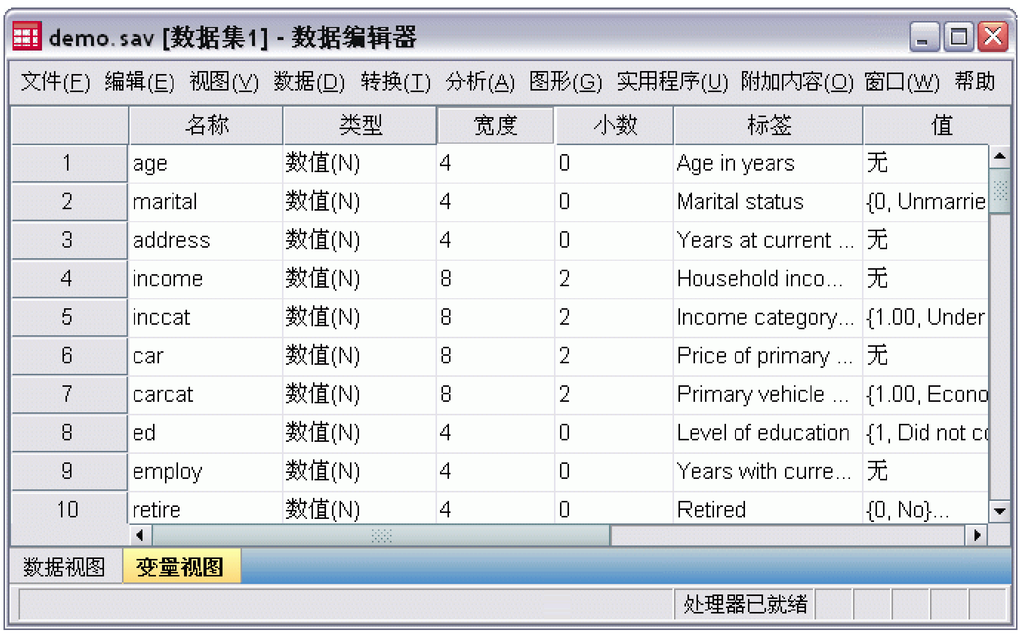
■ 列为变量。每一列代表一个要度量的变量或特征。例如，问卷中的每一项都是一个变量。

■ 单元包含值。每个单元均包含某个个案的某个变量的单个值。单元是个案和变量相交的位置。单元仅包含数据值。与电子表格程序不同，数据编辑器中的单元不能包含公式。

■ 数据文件是矩形的。数据文件的尺寸由个案和变量的个数确定。您可以在任何单元中输入数据。如果在定义的数据文件的界限以外的单元中输入数据，则数据矩形将扩展以包含该单元与文件界限之间的任何行和/或列。数据文件的界限内没有“空”单元。对于数值变量，空白单元会转换成系统缺失值。对于字符串变量，则将空白单元视为有效值。

变量视图

图片 5-2
变量视图



	名称	类型	宽度	小数	标签	值
1	age	数值(N)	4	0	Age in years	无
2	marital	数值(N)	4	0	Marital status	{0, Unmarrie
3	address	数值(N)	4	0	Years at current ...	无
4	income	数值(N)	8	2	Household inco...	无
5	inccat	数值(N)	8	2	Income category...	{1.00, Under
6	car	数值(N)	8	2	Price of primary ...	无
7	carcat	数值(N)	8	2	Primary vehicle ...	{1.00, Econo
8	ed	数值(N)	4	0	Level of education	{1, Did not co
9	employ	数值(N)	4	0	Years with curre...	无
10	retire	数值(N)	4	0	Retired	{0, No}...

“变量视图”包含对数据文件中每个变量的属性的描述。在“变量视图”中：

■ 行为变量。

■ 列为变量属性。

您可以添加或删除变量，以及修改变量属性，包括：

■ 变量名

■ 数据类型

■ 数字位数或字符个数

■ 小数位数

■ 描述性的变量标签和值标签

- 用户定义的缺失值
- 列宽
- 测量级别

当您保存数据文件时，将保存所有这些属性。

除了在“变量视图”中定义变量属性以外，另外还有两种方法可用于定义变量属性：

- “复制数据属性向导”使您能够使用外部 IBM® SPSS® Statistics 数据文件或当前会话中的另一数据集作为模板，用于定义活动数据集中的文件和变量的属性。您也可以将活动数据集中的变量用作活动数据集中其他变量的模板。“复制数据属性”可以在数据编辑器窗口的“数据”菜单上找到。
- “定义变量属性”（也在数据编辑器窗口的“数据”菜单上提供）扫描您的数据并列出任何选定变量的所有唯一数据值、标识未标注的值以及提供自动标注功能。这对于使用数值代码表示类别（例如，0 = Male，1 = Female）的分类变量特别有用。

显示或定义变量属性

▶ 使数据编辑器成为活动窗口。

▶ 双击位于“数据视图”中列顶端的变量名，或单击**变量视图**选项卡。

▶ 要定义新变量，请在任意空白行中输入一个变量名。

▶ 选择您想要定义或修改的属性。

变量名

以下规则适用于变量名：

- 每个变量名必须是唯一的；不允许重复。
- 变量名最多可包含 64 个字节，并且第一个字符必须是字母或字符 @、# 或 $ 之一。后续字符可以是字母、数字、非标点字符和句点 (.) 的任意组合。在代码页模式中，64 个字节在单字节语言（例如英语、法语、德语、西班牙语、意大利语、希伯来语、俄语、希腊语、阿拉伯语和泰语）中通常意味着 64 个字符，在双字节语言（例如日语、中文和韩语）中则为 32 个字符。许多在代码页模式中只占一个字节的字符串在 Unicode 模式中则会占到两个或更多字节。例如，é 以代码页格式是一个字节，但是以 Unicode 格式就是两个字节；因此 résumé 在代码页文件中是六个字节，但在 Unicode 模式中是八个字节。

  注意：（请注意：字母包括书写日常文字所用的任何非标点字符，这些文字要属于平台字符集所支持的语言。）
- 变量名不能包含空格。
- 变量名第一个位置中的 # 字符将变量定义为临时变量。只能使用命令语法创建临时变量。不能在创建新变量的对话框中将变量的第一个字符指定为 #。
- 第一个位置中的 $ 符号表示变量为系统变量。$ 符号不能作为用户定义的变量的第一个字符。
- 可在变量名中使用句点、下划线和字符 $、# 以及 @。例如，A._$@#1 是一个有效的变量名。

- 应避免用句点结束变量名，因为句点可能被解释为命令终止符。只能使用命令语法创建以句点结束的变量。不能在创建新变量的对话框中创建以句点结束的变量。
- 应避免使用下划线结束变量名，因为这样的名称可能与命令和过程自动创建的变量名冲突。
- 不能将保留关键字用作变量名。保留关键字有：ALL、AND、BY、EQ、GE、GT、LE、LT、NE、NOT、OR、TO 和 WITH。
- 可以用任意混合的大小写字符来定义变量名，大小写将为显示目的而保留。
- 当长变量名需要在输出中换行为多行时，会在下划线、句点和内容从小写变为大写的位置进行换行。

变量测量级别

您可以将测量级别指定为刻度（定距或者定比刻度上的数值数据）、有序或名义。名义数据和有序数据可以是字符串（字母数字）或数值。

- **标定**. 当变量值表示不具有内在等级的类别时，该变量可以作为名义变量；例如，雇员任职的公司部门。名义变量的示例包括地区、邮政编码和宗教信仰。
- **有序**. 当变量值表示带有某种内在等级的类别时，该变量可以作为有序变量；例如，从十分不满意到十分满意的服务满意度水平。有序变量的示例包括表示满意度或可信度的态度分数和优先选择评分。
- **刻度**. 当变量值表示带有有意义的度规的已排序类别时，该变量可以作为刻度（连续）变量对待，以便在值之间进行合适的距离比较。刻度变量的示例包括以年为单位的年龄和以千美元为单位的收入。

注意：对于有序字符串变量，将假定字符串值的字母顺序反映了类别的真实顺序。例如，对于具有 low、medium、high 值的字符串变量，类别的顺序将解释为 high、low、medium，这个顺序是错误的。通常，使用数值代码代表有序数据更为可靠。

对于通过转换创建的新数值变量、来自外部源的数据以及在版本 8 之前创建的 IBM® SPSS® Statistics 数据文件，默认测量级别由以下表格中的条件决定。将以条件在表格中的排列顺序对其进行评估。将应用与数据匹配的第一个条件的测量级别。

条件	测量级别
变量的所有值均缺失	名义
格式为美元或定制货币	连续
格式为日期或时间（不包括月份和星期）	连续
变量包含至少一个非整数值	连续
变量包含至少一个负值	连续
变量不包含少于 10,000 的有效值	连续
变量具有 N 个或更多唯一有效值*	连续
变量不包含少于 10 的有效值	连续
变量具有少于 N 个唯一有效值*	名义

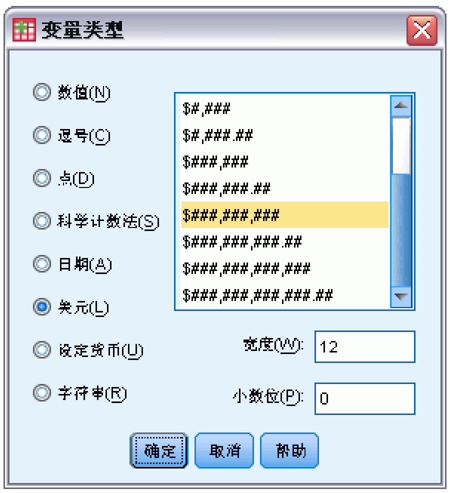
* N 是用户指定的临界值。默认值为 24。

- 您可以在“选项”对话框中更改临界值。
- “数据”菜单中的“定义变量属性”对话框可帮助您指派正确的测量级别。

变量类型

“变量类型”指定每个变量的数据类型。默认情况下，假定所有新变量都为数值变量。您可以使用“变量类型”来更改数据类型。“变量类型”对话框的内容取决于选定的数据类型。对于某些数据类型，存在有关宽度和小数位数的文本框；对于其他数据类型，只需从可滚动的示例列表中选择一种格式即可。

图片 5-3
“变量类型”对话框



可用的数据类型如下：

数值。值为数字的变量。值以标准数值格式显示。数据编辑器接受以标准格式或科学计数法表示的数值。

逗号。变量值显示为每三位用逗号分隔，并用句点作为小数分隔符的数值变量。数据编辑器为逗号变量接受带或不带逗号的数值，或以科学计数法表示的数值。值的小数指示符右侧不能包含逗号。

点。变量值显示为每三位用句点分隔，并带有逗号作为小数分隔符的数值变量。数据编辑器为点变量接受带或不带点的数值，或以科学计数法表示的数值。值的小数指示符右侧不能包含句点。

科学计数法。一个数值变量，它的值以嵌入的 E 以及带符号的 10 次幂指数形式显示。数据编辑器为此类变量接受带或不带指数的数值。指数前面可以加上带符号（可选）的 E 或 D，或只加上符号 — 例如，123、1.23E2、1.23D2、1.23E+2 以及 1.23+2。

日期。一种数值变量，其值以若干种日历-日期或时钟-时间格式中的一种显示。从列表中选择一种格式。输入日期时可以用斜杠、连字符、句号、逗号或空格作为分隔符。两位数年份值的世纪范围由您的“选项”设置确定（从“编辑”菜单中，选择**选项**然后单击**数据**选项卡）。

美元。数值变量，显示时前面带美元符号 ($)，每三位用逗号分隔，并用句点作为小数分隔符。可以输入带有或不带有前导美元符号的数据值。

定制货币。一种数值变量，其值以定制货币格式中的一种显示，定制货币格式是在“选项”对话框的“货币”选项卡中定义的。定义的定制货币字符不能用于数据输入，但显示在数据编辑器中。

字符串。字符串变量的值不是数值，因此不用在计算中。字符串值可以包含任何字符，可包含的最大字符数不超过定义的长度。字符串变量区分大小写字母。此类型又称为字母数值变量。

定义变量类型

▶ 单击您想要定义的变量的类型单元中的按钮。

▶ 在“变量类型”对话框中选择数据类型。

▶ 单击**确定**。

输入格式与显示格式

根据格式，值在“数据视图”中的显示可能与输入的及内部存储的实际值不同。以下是一些通用准则：

- 对于数值、逗号和点格式，您可以输入具有任意小数位数（最多 16 位）的值，整个值会存储在内部。“数据视图”仅显示定义的小数位数，并且对具有更多小数位的值进行舍入。不过，在所有的计算中都使用完整的值。
- 对于字符串变量，所有值都向右填充到最大宽度。对于最大宽度为三的字符串变量，No 值在内部存储为“No ”，并不等于“No”。
- 对于日期格式，您可以使用斜杠、短划线、空格、逗号或句点作为日、月、年值之间的分隔符，并且可以为月值输入数字、三字母缩写或完整名称。以通用格式 dd-mmm-yy 显示日期，其中以短划线作为分隔符并对月份使用三字母缩写。以通用格式 dd/mm/yy 和 mm/dd/yy 显示日期，其中以斜杠作为分隔符并对月份使用数字。日期在内部存储为自 1582 年 10 月 14 日以来的累计秒数。两位数年份的日期世纪范围由您的“选项”设置确定（从“编辑”菜单中，选择**选项**，然后单击**数据**选项卡）。
- 在时间格式中，您可以使用冒号、句号或空格作为小时、分钟和秒数间的分隔符。时间以冒号分隔显示。时间被内部储存为表示时间间隔的秒数。例如，10:00:00 被内部储存为 36000，即 60（秒每分钟）x 60（分钟每小时）x 10（小时）。

变量标签

您可以为描述性变量标签分配最多可达 256 个的字符（在双字节语言中则为 128 个字符）。变量标签可以包含空格和变量名称中所不允许的保留字符。

指定变量标签

▶ 使数据编辑器成为活动窗口。

▶ 双击位于“数据视图”中列顶端的变量名，或单击**变量视图**选项卡。

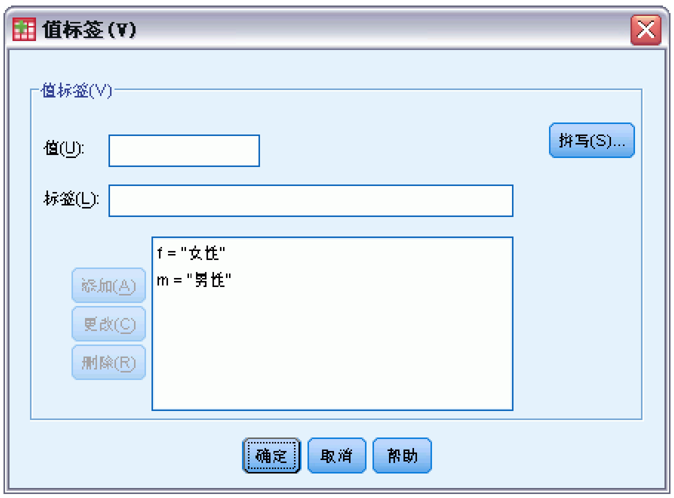
▶ 在变量的标签单元格中，输入描述性变量标签。

值标签

您可以为每个变量值分配描述值标签。当您的数据文件使用数值代码表示非数值类别时（例如：代码 1 和 2 代表 male 和 female），此过程特别有用。

■ 值标签最多可达 120 个字节。

图片 5-4
“值标签”对话框



指定值标签

▶ 单击您想要定义的变量的值单元中的按钮。

▶ 在每个值中，输入值和标签。

▶ 单击**添加**输入值标签。

▶ 单击**确定**。

在标签中插入换行符

若单元格或区域没有足够宽度在一行内显示整个标签，则枢轴表和图表中的变量标签和值标签会自动换行至多行，而如果您想将标签换行至不同点，可以编辑结果，插入手动换行符。您还可以创建变量标签和值标签，这将一直换行至指定点并以多行显示。

▶ 对于变量标签，在数据编辑器的变量视图中选择标签变量单元格。

▶ 对于值标签，在数据编辑器的变量视图中选择值变量单元格，单击单元格中的按钮，选择您想在“值标签”对话框中修改的标签。

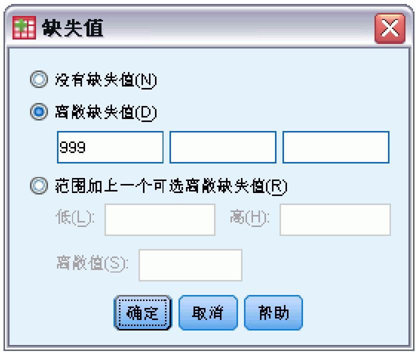
▶ 在标签中您想让标签换行的地方，输入\n。

\n 不在枢轴表或图表中显示；它被翻译为换行字符。

缺失值

缺失值将指定数据值定义为**用户缺失值**。例如，您想要区分因对象拒绝回答问题造成的数据缺失与由于问题不适于该对象而未回答所引起的数据缺失。将指定为用户缺失值的数据值标记为进行特殊处理，并将其从大多数计算中排除。

图片 5-5
“缺失值”对话框



- 您最多可以输入三个离散（单个）缺失值、一个缺失值范围或一个范围加一个离散值。
- 只有数值变量能被指定范围。
- 除非您将其明显定义为缺失，否则所有字符串值包括“零”或空白值都被视为有效。
- 字符串变量的缺失值不得超过八个字节。（字符串变量的宽度定义没有限制，但定义的缺失值不得超过八个字节。）
- 若想将“零”或空白值定义为字符串变量的缺失，则在**离散缺失值**选项下的字段之一输入一个单空格。

定义缺失值

▶ 单击您想要定义的变量的缺失单元中的按钮。

▶ 输入表示缺失数据的值或值范围。

角色

某些对话框支持可用于预先选择分析变量的预定义角色。当打开其中一个对话框时，满足角色要求的变量将自动显示在目标列表中。可用角色包括：

输入。变量将用作输入（例如，预测变量、自变量）。

目标。变量将用作输出或目标（例如，因变量）。

两者。变量将同时用作输入和输出。

无。变量没有角色分配。

分区。变量将用于将数据划分为单独的训练、检验和验证样本。

拆分。包括以便与 IBM® SPSS® Modeler 相互兼容。具有此角色的变量不会在 IBM® SPSS® Statistics 中用作拆分文件变量。

- 默认情况下，为所有变量分配输入角色。这包括外部文件格式的数据和 SPSS Statistics 18 之前版本的数据文件。
- 角色分配只影响支持角色分配的对话框。它对命令语法没有影响。

指定角色

▶ 从变量的角色单元格的列表中选择角色。

列宽

您可以为列宽指定一些字符。也可以通过单击并拖拽列边框来更改数据视图中的列宽。

- 比例字体的列宽是基于平均字符宽度。根据值中使用的字符，在指定宽度内可以显示较多或较少的字符。
- 列宽度只影响数据编辑器中的值显示。更改列宽不会改变变量的已定义宽度。

变量对齐

对齐控制着数据视图中数据值和/或值标签的显示。默认对齐方式为数值变量在右边，字符串变量在左边。此设置只影响数据编辑器中的显示。

将变量定义属性应用于多个变量

在定义了变量的变量定义属性之后,可以复制一个或多个属性并将其应用到一个或多个变量。

基础复制粘贴操作被用于变量定义属性的应用。您可以：

- 复制一个单一属性（例如，值标签）并将其粘贴至一个或多个变量的相同属性单元中。
- 从一个变量中复制所有属性并将其粘贴至一个或多个其他变量。
- 通过一个已复制变量的所有属性创建多个新变量。

将变量定义属性应用于其他变量

从已定义变量中应用单个属性

▶ 在变量视图中，选择您想应用于其他变量的属性单元格。

▶ 从菜单中选择：
编辑 > 粘贴

▶ 选择您想应用其属性的属性单元格。（您可以选择多个目标变量。）

▶ 从菜单中选择：
编辑 > 粘贴

如果您将属性粘贴至空白行，则新变量就与所有属性中除选定属性以外的默认属性一同被创建。

应用已定义变量中的全部属性

▶ 在变量视图中，与您想使用的属性一同选择变量行号。（突出显示整行。）

▶ 从菜单中选择：
编辑 > 粘贴

▶ 与您想使用的属性一同选择变量行号。（您可以选择多个目标变量。）

▶ 从菜单中选择：
编辑 > 粘贴

通过相同属性生成多个新变量

▶ 在变量视图中，单击具有您想用于新变量属性的变量。（突出显示整行。）

▶ 从菜单中选择：
编辑 > 粘贴

▶ 在数据文件中单击最后一个已定义变量下面的空行号码。

▶ 从菜单中选择：
编辑 > 粘贴变量...

▶ 在“粘贴变量”对话框中，输入您想创建的变量数。

▶ 输入新变量的前缀和起始数字。

▶ 单击**确定**。

新变量名将包括指定前缀和由指定数字开始的序列数。

定制变量属性

除了标准变量属性（如值标签、缺失值、测量级别）之外，还可以自己创建定制的变量属性。与标准变量属性一样，这些定制变量属性也随 IBM® SPSS® Statistics 数据文件一同保存。因此，您可以创建识别调查问题响应类型的变量属性（例如，单选、多选、填空）或计算变量使用的公式。

创建定制变量属性

创建新的定制属性：

▶ 在“变量视图”中，从菜单中选择：
数据 > 新建定制属性...

- 将要向其分配新属性的变量拖放到“选定的变量”列表中。
- 为新属性输入名称。属性名必须遵循与变量名相同的规则。
- 为新属性输入一个可选值。如果选择多个变量，则将该值指定给所有选定的变量。可以将此字段保留为空，然后在“变量视图”中输入各变量的值。

图片 5-6
“新建定制属性”对话框



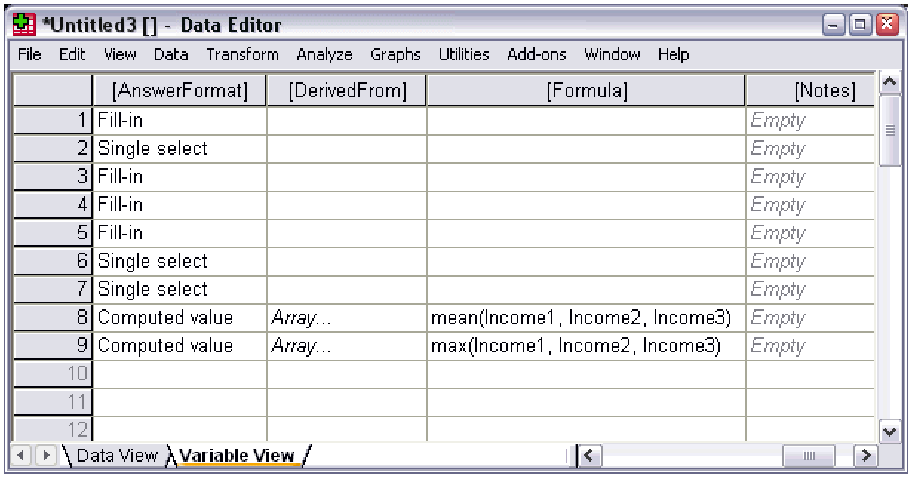
在数据编辑器中显示属性。在数据编辑器的“变量视图”中显示属性。有关控制定制属性显示的信息，请参见下面的显示和编辑定制变量属性。

显示定义的列表属性。显示已为数据集定义的定制属性列表。以美元符号 ($) 开头的属性名称是不能修改的保留属性。

显示和编辑定制变量属性

可以在数据编辑器的“变量视图”中显示和编辑定制变量属性。

图片 5-7
“变量视图”中显示的定制变量属性



- 定制变量属性的名称用方括号括起。
- 以美元符号开头的属性名是保留名称，不能修改这些名称。
- 空单元格表示该变量没有属性；单元格中显示为 Empty 的文本表示该变量具有属性，但还没有为该变量的属性赋值。在单元格中输入文本后，该变量即拥有了具有您所输入的值的属性。
- 在单元格中显示的数组... 表示此属性是属性数组，即包含多个值的属性。单击单元格中的按钮以显示值列表。

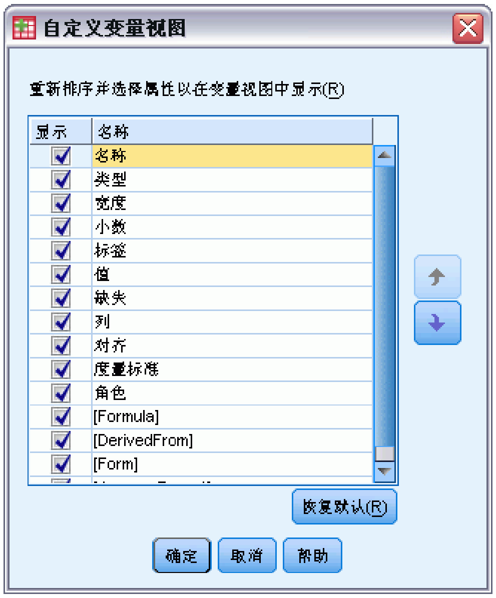
显示和编辑定制变量属性

▶ 在“变量视图”中，从菜单中选择：

视图 > 自定义变量视图...

▶ 选择（选中）要显示的定制变量属性。（定制变量属性用方括号括起。）

图片 5-8
自定义变量视图

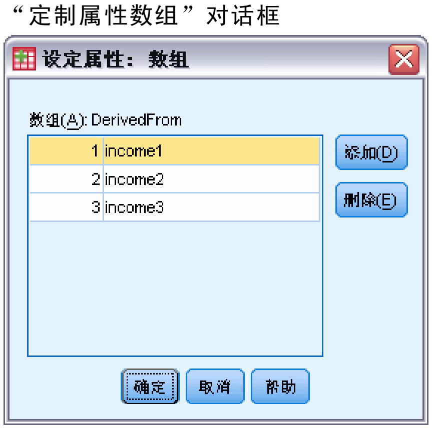


一旦在“变量视图”中显示属性，您即可直接在数据编辑器中对其进行编辑。

变量属性数组

在“变量视图”的定制变量属性单元格中或者在“定义变量属性”的“定制变量属性”对话框中显示的文本数组...表示：该属性是一个**属性数组**，即一个包含多个值的属性。例如，可以有一个属性数组，它标识用于计算派生变量的所有源变量。单击单元格中的按钮可以显示和编辑值列表。

图片 5-9
“定制属性数组”对话框



自定义变量视图

您可以使用自定义变量视图控制变量视图中显示的属性（例如，名称、类型、标签）及其显示顺序。

- 任何与数据库有关的定制变量属性都以方括号括起。
- 自定义显示设置保存在 IBM® SPSS® Statistics 数据文件中。
- 您还可以控制变量视图中的默认显示和属性顺序。

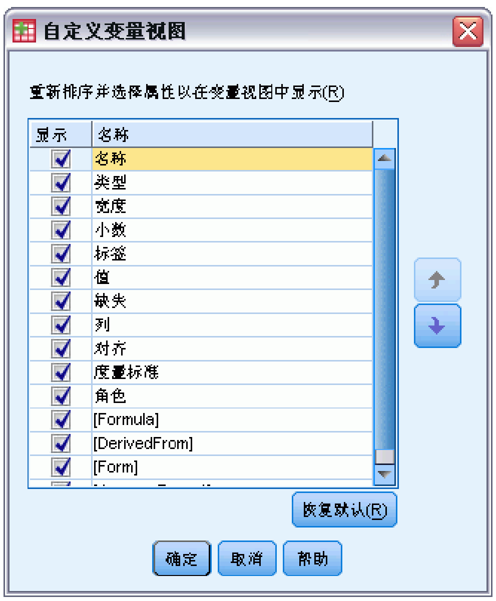
自定义变量视图

▶ 在“变量视图”中，从菜单中选择：
视图 > 自定义变量视图...

▶ 选择（选中）要显示的变量属性。

▶ 使用向上和向下箭头按钮更改属性的显示顺序。

图片 5-10
“自定义变量视图”对话框



恢复默认。应用默认显示与顺序设置。

拼写检查

变量和值标签

检查变量标签与值标签的拼写：

▶ 在“数据编辑器”窗口中，选择“变量视图”选项卡。

▶ 右键单击标签或值列，并从上下文菜单中选择：
拼写

或

▶ 在“变量视图”中，从菜单中选择：
实用程序 > 拼写

或

▶ 在对话框中单击**拼写**。（这将值标签的拼写检查限制为特定变量。）

拼写检查仅限于数据编辑器的变量视图中的变量标签和值标签。

字符串数据值

检查字符串数据值的拼写：

▶ 在“数据编辑器”窗口中，选择“数据视图”选项卡。

▶ （可选）选择一个或多个待检查的变量（列）。要选择变量，请单击列顶部的变量名。

▶ 从菜单中选择：
实用程序 > 拼写

- 如果在“数据视图”中没有选定的变量，则会选中所有字符串变量。
- 如果数据集不包含字符串变量，或选定变量均非字符串变量，则“实用程序”上的“拼写”选项被禁用。

输入数据

在数据视图中，您可以在数据编辑器里直接输入数据。您可以以任何顺序输入数据。您可以在选定区域或个别单元格中，按个案或按变量输入数据。

- 突出显示活动单元格。
- 在数据编辑器左上角显示变量名称和活动单元格的行号。
- 当您选择一个单元格输入数据值时，该值显示在数据编辑器顶端的单元格编辑器中。
- 当按下回车键或选择另一个单元格时，数据值方被保存。
- 欲输入除简单数值数据外的任何其他数据时，您必须先定义变量类型。

如果在空列中输入一个值，数据编辑器将自动创建一个新变量并赋予一个变量名。

输入数值数据

▶ 在数据视图中选择一个单元格。

▶ 输入数据值。（值被显示在数据编辑器顶端的单元格编辑器中。）

▶ 按回车键或选择另一个单元格，即可保存值。

输入非数值数据

▶ 双击位于“数据视图”中列顶端的变量名，或单击**变量视图**选项卡。

▶ 单击变量的类型单元格。

▶ 在“变量类型”对话框中选择数据类型。

▶ 单击**确定**。

▶ 双击行号或单击**数据视图**选项卡。

▶ 在新近定义的变量列中输入数据。

使用值标签进行数据输入

▶ 若值标签没有显示在当前数据视图中，就从菜单中选择：
视图 > 值标签

▶ 单击您想输入值的单元格。

▶ 从下拉列表中选择值标签。

值被输入后，值标签就显示在单元格里。

注意：此过程仅当您已对变量的值标签定义时才会运行。

数据编辑器中的数据值限制

既定变量类型和宽度决定了在数据视图单元格中可以输入的值类型。

- 如果输入了一个既定变量类型不允许的字符，则该字符无法输入。
- 对于字符串变量，不允许字符超出既定宽度。
- 对于数值变量，可以输入超出既定宽度的整数值，但数据编辑器既显示科学计数法又显示一部分值，其后用省略号（...）指明该值已超出既定宽度。更改变量的既定宽度即可在单元格中显示该值。

  注意：更改列宽不会影响变量宽度。

编辑数据

通过数据编辑器，您可以在数据视图中以多种方法修改数据值。您可以：

- 更改数据值
- 剪切、复制并粘贴数据值
- 添加删除个案
- 添加删除变量
- 更改变量顺序

替换或修改数据值

删除原有值，输入新值

▶ 在数据视图中双击单元格。（单元格值显示在单元格编辑器中。）

▶ 在单元格或单元格编辑器中直接编辑值。

▶ 按回车键或选择另一个单元格保存新值。

剪切、复制并粘贴数据值

在数据编辑器中您可以对单个单元格值或组值进行剪切、复制和粘贴。您可以：

- 将一个单一单元格值移动或复制到另一个单元格
- 将一个单一单元格值移动或复制到一组单元格
- 将单个个案（行）的值移动或复制到多个个案
- 将单个变量（列）的值移动或复制到多个变量
- 将一组单元格值移动或复制到另一组单元格

数据编辑器中已粘贴值的数据转换

若源单元格与目标单元格的既定变量类型不同，数据编辑器会试图将该值转换。如果无法转换，将在目标单元格中插入系统缺失值。

将数值或日期转换为字符串。数值（例如，数值、美元符号、点或逗号）和日期格式若被粘贴至字符串变量单元格，则将被转换为字符串。在单元格中显示时，字符串值为数值型值。例如，对于一个美元符号格式的变量，显示的美元符号图标则变成字符串值的一部分。超出既定字符串值宽度的值将被截去。

将字符串转换为数值或日期。包含被目标单元格的数值或日期格式所接受字符的字符串值将被转换为同等的数值或日期值。例如，如果目标单元格的格式类型为日-月-年格式中的一种，则字符串 25/12/91 将被转换为有效日期，但如果目标单元格的格式类型为月-日-年格式中的一种，则该值将被转换为系统缺失值。

将日期转换为数值。如果目标单元格为数值格式中的一种，则日期和时间值将被转换为秒数（例如，数值、美元符号、点或逗号）。因为日期从 1582 年 10 月 14 日开始被内部储存为秒数，且将日期转换为数值可以产生一些非常大的数字。例如，日期 10/29/91 被转换为数值 12,908,073,600。

将数值转换为日期或时间。如果该值表示能产生有效日期或时间的秒数，则该数值将被转换为日期或时间。对于日期，小于 86,400 的数值将被转换为系统缺失值。

插入新个案

在空行的单元格值输入数据会自动创建一个新个案。数据编辑器将为所有个案中的所有其他变量插入系统缺失值。若在新个案和已有个案间存在任何空行，则该空行将成为具有用于所有变量的系统缺失值的新个案。您也可以在已有个案之间插入新个案。

在已有个案之间插入新个案

- 在数据视图中，在想要插入新个案地方的下方个案（行）中任选一个单元格。
- 从菜单中选择：
  编辑 > 插入个案

  在个案中插入新行，而所有变量都被赋予系统缺失值。

插入新变量

在数据视图中的空列中或在变量视图中输入数据将自动创建一个具有默认变量名（前缀 var 和序列数）和默认数据格式类型（数值）的新变量。数据编辑器将为新变量插入用于所有个案的系统缺失值。若在数据视图中存在任何空列或在变量视图中的新变量和已有变量间存在任何空行，该行或列都将成为具有用于所有个案的系统缺失值的新变量。您也可以在已有变量之间插入新变量。

在已有变量之间插入新变量

- 在（数据视图）变量右侧或想要插入新变量地方的下方（变量视图）任选一个单元格。
- 从菜单中选择：
  编辑 > 插入变量

  新变量被插入了用于所有个案的系统缺失值。

移动变量

- 要选取变量，请单击数据视图中的变量名或变量视图中的变量行号。
- 将变量拖放到新位置中。
- 如果您想在两个已有变量之间放置变量，则：在数据视图中，将值放置在您想放置变量地方的右侧变量列中，或者在变量视图中将值放置在您想放置变量地方的下方变量行中。

更改数据类型

您随时可以通过使用变量视图中的“变量类型”对话框来更改变量的数据类型。数据编辑器将试图将已有值转换为新类型。若无法转换，则赋予其系统缺失值。该转换规则与将不同格式类型的数据值粘贴至变量的规则相同。若数据格式的更改会引起缺失值规格或值标签的丢失，数据编辑器将显示提示对话框并询问您是否想要继续此更改或将其取消。

查找个案、变量或插补

在数据编辑器中进入“转到”对话框查找指定个案（行）号或变量名。

个案

▶ 对于个案，从菜单中选择：
编辑 > 转至个案...

▶ 在数据视图中输入一个表示当前行号的整数值。

注意：个别个案的当前行号可以因排序和其他操作而更改。

变量

▶ 对于变量，从菜单中选择：
编辑 > 转向变量...

▶ 输入变量名或从下拉列表中选择变量。

图片 5-11
“转到”对话框



插补

▶ 从菜单中选择：
编辑 > 转至归因…

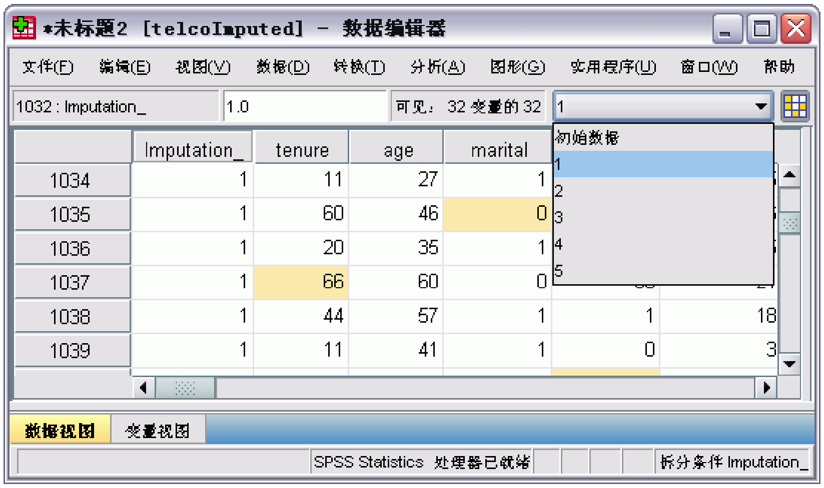
▶ 从下拉列表中选择插补（或原始数据）。

图片 5-12
“转到”对话框



或者，也可从“数据编辑器”的数据视图上编辑栏的下拉列表中选择插补。

图片 5-13
“数据编辑器”（插补标记打开）



在选择插补时相对个案位置被保留。例如，如果在源数据集中有 1000 个个案，个案 1034，即第 1 个插补中的第 34 个个案，显示在网格顶部。如果您从下拉列表中选择插补 **2**，则个案 2034，即第 2 个插补中的第 34 个个案，将显示在网格顶部。如果您从下拉列表中选择**原始数据**，则个案 34 将显示在网格顶部。在两个插补之间浏览时，列位置也将被保留，这样可以方便地进行值比较。

查找并替换数据和属性值

在数据视图中查找和/或替换数据值或者在变量视图中查找和/或替换属性值：

▶ 在列中单击想要搜索的单元格。（禁止对单列的值进行查找和替换。）

▶ 从菜单中选择：

编辑 > 查找

或

编辑 > 替换

数据视图

- 不得在数据视图中搜索。搜索向导一般为不可按。
- 对于日期和时间，将搜索“数据视图”中显示的格式化值。例如，显示为 10/28/2007 的日期若以 10-28-2007 日期的形式就无法被找到。
- 对于其他数值变量，**包含**、**开始以**和**结束以**将搜索格式化值。例如，使用**开始以**选项，搜索值为 $123 的美元格式变量将找到 $123.00 和 $123.40，但找不到 $1,234。使用**整个单元格**选项，搜索值可以是已格式化的或未格式化的（简单 F 数值格式），但只匹配精确数值（根据“数据编辑器”中显示的精确度）。
- 若值标签显示为选定变量列，则标签文本而不是潜在数据值被找到，且您不能替换该标签文本。

变量视图

- 查找仅用于名称、标签、值、丢失以及定制变量属性列。
- 替换仅用于标签、值以及定制变量属性列。
- 在值（值标签）列中，搜索字符串可与数据值或值标签匹配。

  注意：对数据值的替换将删除任何与该值有关的原有值标签。

数据编辑器中的个案选择状态

如果您选择了一个个案子集，但尚未抛弃未选择的个案，则数据编辑器中用穿过行号的对角线（斜杠）来标记未选择的个案。

图片 5-14
数据编辑器中的已过滤个案



数据编辑器显示选项

视图菜单为数据编辑器提供了多个显示选项：

字体。此选项控制数据显示的字体特征。

网格线。此选项可在网格线的显示之间转换。

值标签。此选项可在实际数据值的显示与用户定义的描述性值标签之间转换。此选项只在数据视图中可用。

使用多个视图

在数据视图中，您可以通过使用位于水平横轴下端、垂直数轴右侧的拆分器创建多个视图（窗格）。

您也可以使用窗口菜单插入并删除窗格拆分。插入拆分：

▶ 在数据视图中，从菜单中选择：
窗口 > 拆分

拆分器被插入到上方及选定单元格的左侧。

- 若左上角的单元格被选中，则拆分器的插入将当前视图水平垂直大致平分。
- 若除第一列中的顶端单元格外的任何单元格被选中，则水平窗格拆分器被插入于选定单元格的上方。
- 若除顶端行中的第一个单元格外的任何单元格被选中，则垂直窗格拆分器被插入于选定单元格的左侧。

数据编辑器打印

数据文件将按照其在屏幕上的显示打印。

- 当前显示视图中的信息被打印。在数据视图中，数据被打印。在变量视图中，数据定义信息被打印。
- 若网格线在当前选中视图中有所显示，就会被打印。
- 若数据视图中的值标签在当前有所显示，就会被打印。此外，实际数据值被打印。

使用数据编辑器中的视图菜单显示或隐藏网格线并在数据值和值标签的显示间转换。

打印数据编辑器目录

▶ 使数据编辑器成为活动窗口。

▶ 单击要打印视图的选项卡。

▶ 从菜单中选择：
文件 > 打印...

章

6

使用多数据源

从版本 14.0 开始，可以同时打开多个数据源，从而使得以下操作更加轻松：

- 在数据源之间来回切换。
- 比较不同数据源的内容。
- 在数据源之间复制和粘贴。
- 创建个案和/或变量的多个子集以进行分析。
- 在不事先保存每个数据源的情况下，合并来自各种数据格式（例如：电子表格、数据库、文本数据）的多个数据源。

多数据源的基本处理

图片 6-1
同时打开两个数据源

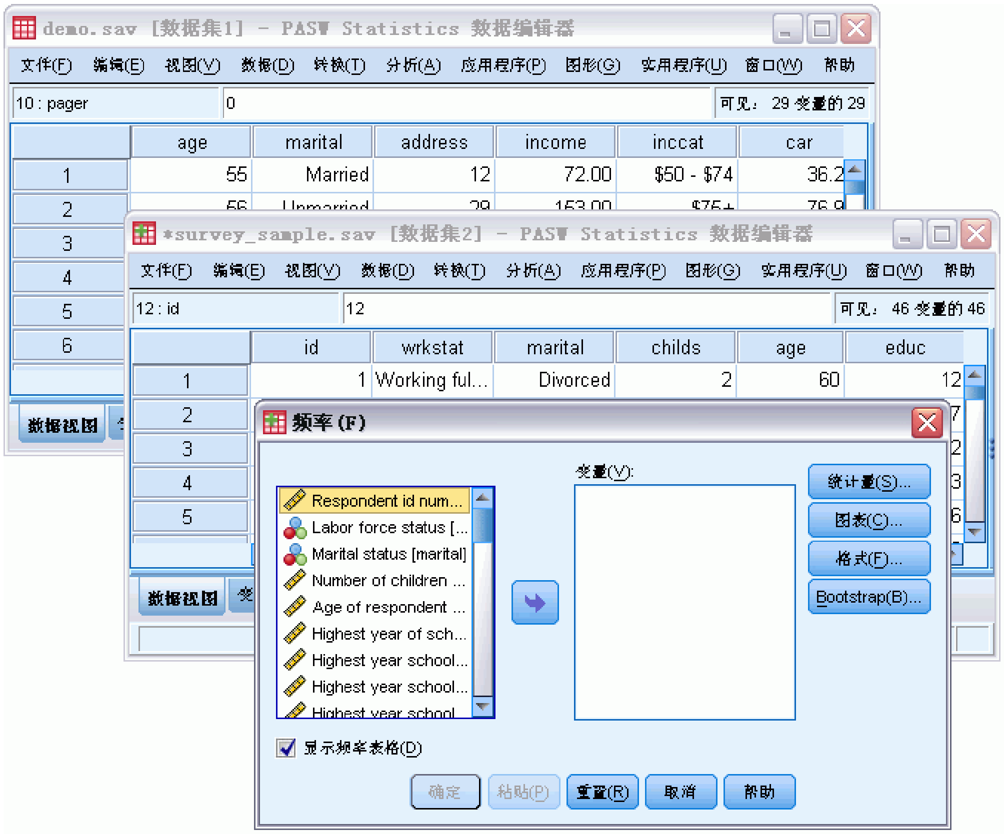


缺省情况下，所打开的每个数据源都显示在新的数据编辑器窗口中。（请参见 一般选项 以获取有关更改缺省行为，以便在单个数据编辑器窗口中一次只能显示一个数据集的信息。）

- 以前打开的任何数据源都保持打开状态，并可供未来使用。
- 第一次打开数据源时，它自动成为**活动数据集**。
- 只需通过单击要使用的数据源的数据编辑器窗口中任意位置，或从“窗口”菜单中选择该数据源的数据编辑器窗口，就可以更改活动数据集。

■ 只有活动数据集中的变量才能用于分析。

图片 6-2
包含活动数据集中的变量的变量列表



■ 访问数据的任何对话框打开（包括显示变量列表的所有对话框）时，不能更改活动数据集。

■ 在会话过程中，必须至少打开一个数据编辑器窗口。关闭最后一个打开的数据编辑器窗口时，IBM® SPSS® Statistics 将自动关闭，提示首先保存更改。

使用命令语法中的多个数据集

如果使用打开数据源的命令语法（例如 GET FILE、GET DATA），则需要使用 DATASET NAME 命令显式地命名每个数据集，以便同时打开多个数据源。

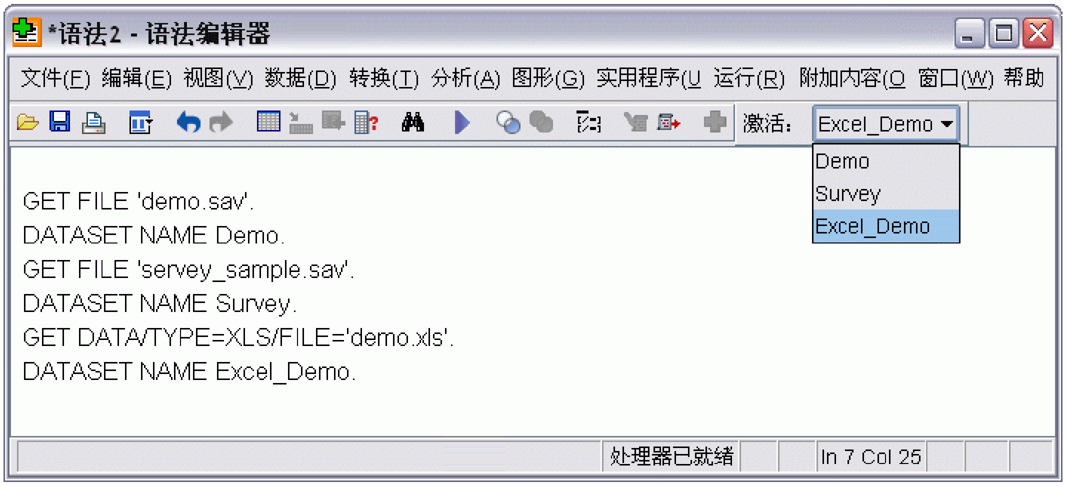
在使用命令语法时，活动数据集名称显示在语法窗口的工具栏上。所有以下操作均可更改活动数据集：

■ 使用 DATASET ACTIVATE 命令。

■ 单击数据集的“数据编辑器”窗口中的任意位置。

■ 从语法窗口中选择数据集名。

图片 6-3
打开在语法窗口工具栏上显示的数据集



在数据集之间复制和粘贴信息

可以使用与在单个数据文件中复制和粘贴信息基本相同的方法，将数据和变量定义属性从一个数据集复制到另一数据集。

- 在“数据视图”中复制和粘贴所选数据单元将仅粘贴数据值，而不包括变量定义属性。
- 通过选择在列顶部的变量名称来复制和粘贴“数据视图”中的整个变量，会粘贴该变量的所有数据和所有数据定义属性。
- 在“变量视图”中复制和粘贴变量定义属性或整个变量将粘贴所选属性（或整个变量定义），但不粘贴任何数据值。

重命名数据集

通过菜单和对话框打开数据源时，向每个数据源都自动分配 DataSetn 的数据集名，其中 n 是序数整数值；而使用命令语法打开数据源时，不会分配数据集名，除非用 DATASET NAME 显式地指定名称。提供更具描述性的数据集名：

▶ 从要更改名称的数据集的“数据编辑器”窗口中的菜单，选择：

文件 > 重命名数据集...

▶ 输入符合变量命名规则的新数据集名。

不显示多个数据集

如果您希望一次只提供一个数据集，并取消多个数据集功能：

▶ 从菜单中选择：

编辑 > 选项...

▶ 单击**常规**选项卡。

选择（选中）**一次只能打开一个数据集**。

章

7

数据准备

一旦您打开了某个数据文件，或在数据编辑器中输入了数据，您就可以开始创建报表、图表和分析，而无需任何附加的预备工作。不过，您可能会觉得某些附加的数据准备功能很有用，包括可以：

- 指定可以描述数据和确定应如何处理特定值的变量属性。
- 确定可能包含重复信息的个案，并将这些个案从分析中排除或从数据文件中删除。
- 创建新变量，这些新变量所具有的一些不同类别代表了具有大量可能值的变量中的值范围。

变量属性

在数据编辑器的数据视图中输入的数据或从外部文件格式（例如，Excel 电子表格或文本数据文件）读入到的数据，缺乏某些您可能觉得非常有用的变量属性，包括：

- 数值代码的描述值标签的定义（例如：0 = Male 和 1 = Female）。
- 缺失值代码的标识（例如：99 = Not applicable）。
- 指定测量级别（名义、有序或刻度）。

所有这些变量属性（及其他属性）均可以在数据编辑器的“变量视图”中指定。还有多个实用程序可在此过程中对您有所帮助：

- **定义变量属性**可帮助您定义描述值标签和缺失值。对于带有用作类别值的数值代码的分类数据，这就特别有用。
- **设置未知测量级别**标识未定义测量级别的变量（字段），并能够为这些变量设置测量级别。这对于测量级别能够影响结果或决定可用功能的过程非常重要。
- **复制数据属性**使您能够使用现有的 IBM® SPSS® Statistics 数据文件作为当前数据文件中的文件和变量属性的模板。如果您经常使用包含相似内容的外部格式数据文件（如 Excel 格式的月度报表），这就特别有用。

定义变量属性

“定义变量属性”在为变量指定属性，包括为类别（名义、有序）变量创建描述性值标签的过程中为您提供帮助。定义变量属性：

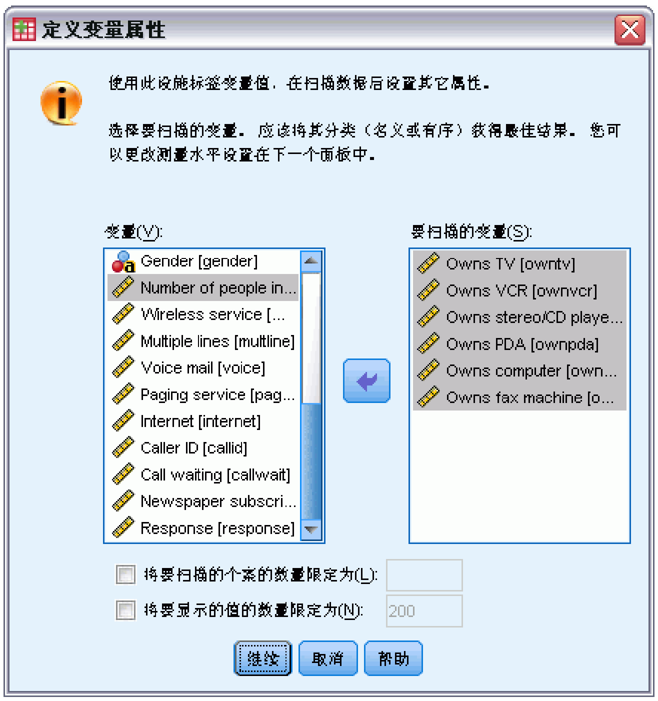
- 扫描实际数据值并为每个选定的变量列出所有的唯一数据值。
- 识别未标注的值并提供“自动标注”功能。
- 提供将已定义的值标签和其他属性从另一变量复制到选定的变量，或是从选定的变量复制到多个其他变量的功能。

注意：要在不先扫描个案的情况下使用“定义变量属性”，则对要扫描的个案数输入 0。

定义变量属性

▶ 从菜单中选择：
数据 > 定义变量属性...

图片 7-1
用于选择要定义的变量的初始对话框



▶ 选择要为其创建值标签或要定义或更改其他变量属性（例如缺失值或描述性变量标签）的数值或字符串变量。

▶ 指定要进行扫描以生成唯一值列表的个案数。这对具有大量个案的数据文件特别有用，因为这种情况下扫描完整数据文件可能需要很长时间。

▶ 指定要显示的唯一值数目的上限。这主要为了防止出现为刻度（连续区间、比率）变量列出上百、上千甚至上百万个值的情况。

▶ 单击**继续**以打开“定义变量属性”主对话框。

▶ 选择要为其创建值标签或要定义或更改其他变量属性的变量。

▶ 为在“值标签”网格中显示的任何未标记的值输入标签文本。

▶ 如果存在要为其创建值标签的值，但是并未显示这些值，可在位于最后一个扫描值之下的值列中输入值。

▶ 对要为其创建值标签的每个列出的变量重复该过程。

▶ 单击**确定**以应用值标签和其他变量属性。

定义值标签和其他变量属性

图片 7-2
定义变量属性，主对话框



“定义变量属性”主对话框提供已扫描变量的下列信息：

已扫描的变量列表。对每一个已扫描的变量，未标记 (U.) 列中的复选标记指示该变量包含未指定值标签的值。

要对变量列表排序以在列表顶部显示所有含未标记值的变量：

▶ 单击“已扫描的变量列表”下的未标记列标题。

您也可单击“已扫描的变量列表”下相应的列标题，按变量名和测量级别进行排序。

值标签网格

- **标签**。显示任何已定义的值标签。您可在该列添加或更改标签。
- **值**。每个选定的变量的唯一值。该唯一值列表基于已扫描的个案数。例如，如果您仅扫描了数据文件中的前 100 个个案，那么列表仅反映这些个案中的唯一值。如果数据文件已按变量排序，而您想要为该变量指定值标签，列表中显示的唯一值可能远远少于数据中实际有的唯一值。
- **计数**。每一个值出现在已扫描个案中的次数。

- **缺失。**定义为代表缺失数据的值。您可以单击复选框来更改类别的缺失值指定。复选标记指示该类别已被定义为用户缺失类别。如果某个变量具有定义为用户缺失值的值范围（例如，90-99），则不能使用“定义变量属性”为该变量添加或删除缺失值类别。您可使用数据编辑器中的“变量视图”为具有缺失值范围的变量修改缺失值类别。
- **已更改。**指示您已添加或更改值标签。

注意：如果在初始对话框中指定要扫描的个案数为 0，除预先存在的值标签和/或为选定变量定义的缺失值类别之外，“值标签”网格初始情况下将为空白。此外，测量级别的**建议**按钮将被禁用。

测量级别。值标签主要对分类（标定和有序）变量很有用，某些过程区别对待分类变量和尺度变量；因此，指定正确的测量级别有时很重要。但是，在缺省情况下，所有新的数值变量都将被指定该刻度测量级别。所以，许多事实上为分类变量的变量可能最初显示为刻度变量。

如果您不确定要将哪一种测量级别指定给变量，单击**建议**。

角色。 某些对话框支持基于定义的角色预先选择分析变量的功能。

复制属性。您可以将值标签和其他变量属性从另一变量复制到当前选定的变量，或是从当前选定的变量复制到一个或多个其他变量。

未标记的值。要为未标记的值自动创建标签，请单击**自动标记**。

变量标签和显示格式

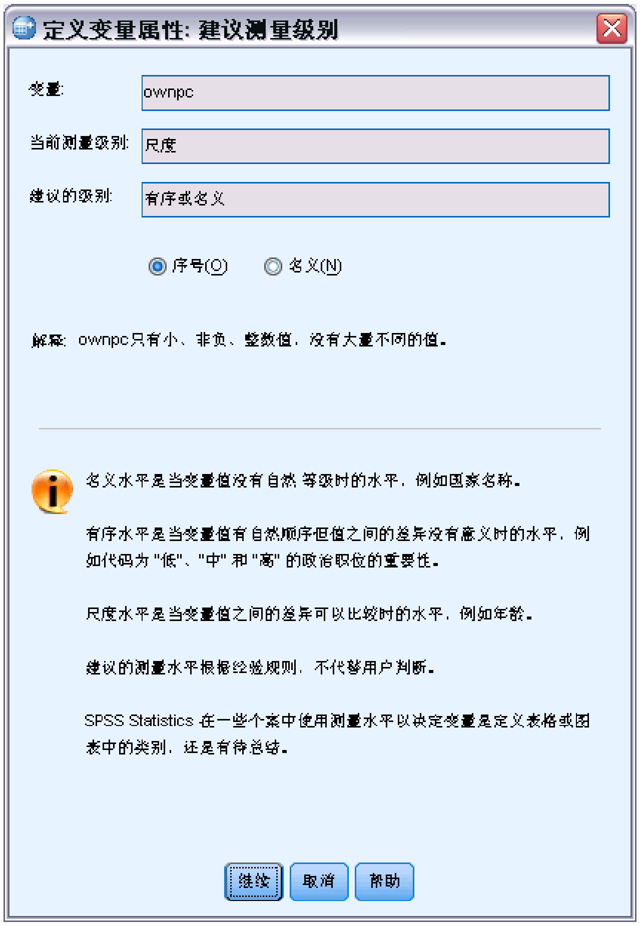
您可更改描述性变量标签和显示格式。

- 您不能更改变量的基本类型（字符串或数值）。
- 对字符串变量，您只可更改变量标签，而不能更改显示格式。
- 对数值变量，您可以更改数值类型（例如数值、日期、美元或定制货币）、宽度（最大位数，包括小数和/或分组指示符）和小数位数。
- 对数值日期格式，您可选择一个特定日期格式（例如 dd-mm-yyyy、mm/dd/yy 和 yyyyddd）
- 对数值定制格式，您可选择五个定制货币格式中的一个（CCA 到 CCE）。
- 如果指定宽度小于预先存在的已定义值标签或缺失值类别的扫描值或显示值的宽度，在值列中将显示星号。
- 预先存在的已定义值标签或缺失值类别的扫描值或显示值对选定的显示格式类型无效，则显示一个句点 (.)。例如，小于 86,400 的内部数值对日期格式变量是无效的。

指定测量级别

当您对“定义变量属性”主对话框中的测量级别单击**建议**时，将基于已扫描的个案和定义的值标签计算当前变量，打开的“建议测量级别”对话框将建议一个测量级别。“说明”区域提供了用于提供测量级别建议的标准的简要描述。

图片 7-3
“建议测量级别”对话框



注意：测量级别计算中不包含定义为代表缺失值的值。例如，测量级别建议的说明可能表示该建议的部分依据是变量不包含负值这一事实，然而，事实上变量可能包含负值，但是这些负值已被定义为缺失值。

▶ 单击**继续**以接受建议的测量级别，或单击**取消**，不更改测量级别。

设定变量属性

使用“定义变量属性”中的**属性**按钮可打开“定制变量属性”对话框。除标准变量属性（例如，值标签、缺失值和测量级别）外，还可以创建自己的定制变量属性。与标准变量属性一样，这些定制变量属性也随 IBM® SPSS® Statistics 数据文件一同保存。

图片 7-4
设定变量属性



名称。属性名必须遵循与变量名相同的规则。

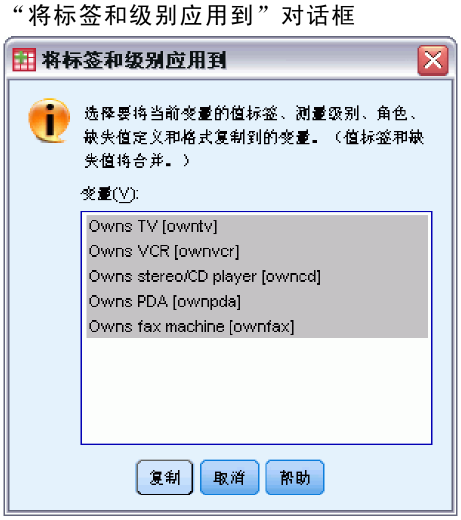
值。指定给选定变量的属性的值。

- 以美元符号开头的属性名是保留名称，不能修改这些名称。您可以单击所需单元格中的按钮以查看保留属性的内容。
- 在“值”单元格中显示的数组... 表示此属性是**属性数组**，即包含多个值的属性。单击单元格中的按钮以显示值列表。

复制变量属性

当您单击“定义变量属性”主对话框中的**从另一个变量或到其他变量**时，将显示“应用标签和水平”对话框。将显示与当前变量类型（数值或字符串）匹配的所有已扫描的变量。对字符串变量，定义的宽度也必须匹配。

图片 7-5
“将标签和级别应用到”对话框



▶ 选择要从中复制值标签和其他变量属性（变量标签除外）的单个变量。

或

▶ 选择要将值标签和其他变量属性复制到其中的一个或多个变量。

▶ 单击**复制**以复制值标签和测量级别。

- 此操作不替换目标变量的现有值标签和缺失值类别。
- 没有为目标变量定义的值标签和缺失值类别将被添加至目标变量的值标签和缺失值类别集中。
- 此操作总是替换目标变量的测量级别。
- 此操作总是替换目标变量的角色。
- 如果源或目标变量中的任何一个含有定义的缺失值范围，那么将不复制缺失值定义。

为测量级别未知的变量设置测量级别

对于某些过程，测量级别可能影响结果，或决定哪些功能可用，并且您将无法访问这些过程的对话框，除非所有变量均已定义测量级别。“设置未知测量级别”对话框允许您为任何测量级别未知的变量定义测量级别，而无需执行数据传递（对于大型数据文件这可能非常耗时）。

在特定情况下，文件中的某些或全部数值变量（字段）的测量级别可能未知。这些情况包括：

- 在首次数据传递之前，来自 Excel 95 或更高版本文件、文本数据文件或数据库源的数值变量。
- 在这些变量创建之后但在首次数据传递之前，通过转换命令创建的新数值变量。

这些情况主要适用于通过命令语法读取数据或创建新变量。读取数据和创建新转换变量的对话框会自动执行数据传输，以根据默认测量级别规则来设置测量级别。

为测量级别未知的变量设置测量级别

▶ 在为过程显示的警报对话框中，单击**手动指定**。

或

▶ 从菜单中选择：

数据 > 设置未知测量级别

图片 7-6
“设置未知测量级别”对话框



▶ 将变量（字段）从源列表移动到适当的测量级别目标列表。

- **标定**. 当变量值表示不具有内在等级的类别时，该变量可以作为名义变量；例如，雇员任职的公司部门。名义变量的示例包括地区、邮政编码和宗教信仰。
- **有序**. 当变量值表示带有某种内在等级的类别时，该变量可以作为有序变量；例如，从十分不满意到十分满意的服务满意度水平。有序变量的示例包括表示满意度或可信度的态度分数和优先选择评分。
- **连续**. 当变量值表示带有有意义的度规的已排序类别时，该变量可以作为刻度（连续）变量对待，以便在值之间进行合适的距离比较。刻度变量的示例包括以年为单位的年龄和以千美元为单位的收入。

多重响应集

定制表和图表生成器支持一种称为**多重响应集**的特殊“变量”。多重响应集不是通常意义上真正的“变量”。多重响应集不显示在数据编辑器中，也不能由其他过程识别。多重响应集使用多个变量记录对问题的答复，其中响应者可以给出多个答案。多重响应集以与分类变量相同的方式对待，可以对分类变量执行的大多数操作也可以对多重响应集执行。

多重响应集根据数据文件中的多个变量进行构造。多重响应集是数据文件中的特殊结构。可以在 IBM® SPSS® Statistics 数据文件中定义和保存多重响应集，但不能从其他文件格式导入或向其导出多重响应集。可以使用“复制数据属性”从其他 SPSS Statistics 数据文件复制多重响应集，从“数据编辑器”窗口的“数据”菜单可以访问该功能。

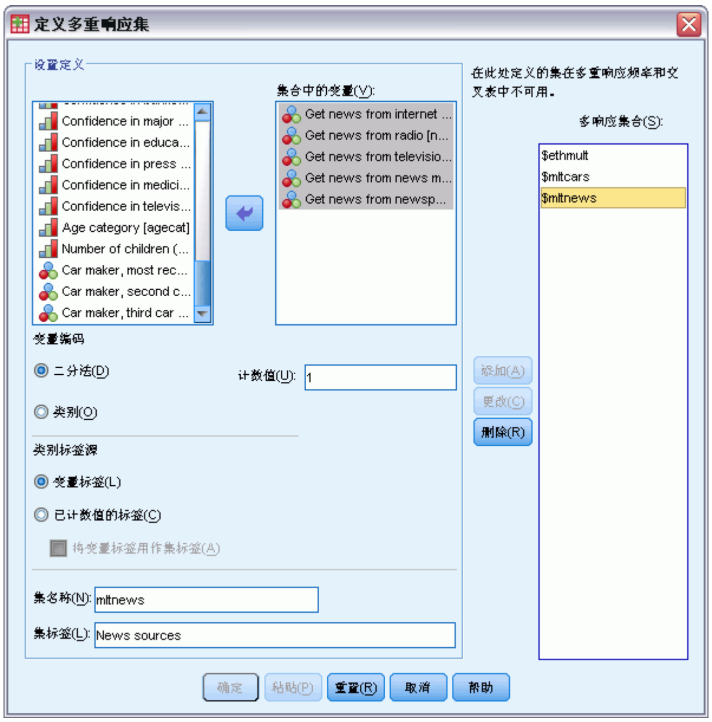
定义多重响应集

定义多重响应集：

▶ 从菜单中选择：

数据 > 定义多重响应集...

图片 7-7
“定义多重响应集”对话框



▶ 选择两个或更多个变量。如果变量以二分法进行编码，则请指明要计数的值。

▶ 为每个多重响应集输入唯一名称。名称最多可以包含 63 个字节。将自动向集名称的开头添加一个美元符号。

▶ 为集合输入描述性标签。（此为可选。）

▶ 单击**添加**将多重响应集添加到定义的集的列表中。

二分

多二分集通常由多个二分变量组成：仅有两个可能值（是/否、存在/不存在、选中/未选中性质）的变量。尽管变量可能不是严格二分的，但集中的所有变量都以相同方式进行编码，且“已计算的值”表示肯定/存在/选中条件。

例如，调查提出以下问题：“对于新闻，您依赖于以下哪些来源？”并提供了五个可能的回答。回答者可以通过选中每个选择旁的框来进行多项选择。五个回答成为数据文件中的五个变量，代码 0 表示否（未选中），代码 1 表示是（选中）。在多二分集中，“已计算的值”为 1。

样本数据文件 survey_sample.sav 已有三个已定义的多重响应集。$mltnews 是多二分集。

▶ 选择（单击）多重响应集列表中的 $mltnews。

此操作会显示用于定义此多重响应集的变量和设置。

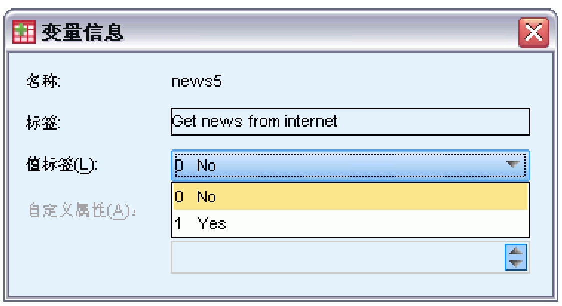
- “集合中的变量”列表显示用于构成多重响应集的五个变量。
- “变量编码”组表示变量是二分的。
- “已计算的值”为 1。

▶ 在“集合中的变量”列表中选择（单击）其中一个变量。

▶ 在弹出式上下文菜单中右键单击变量并选择**变量信息**。

▶ 在“变量信息”窗口中，单击“值标签”下拉列表上的箭头，以显示已定义的值标签的完整列表。

图片 7-8
多二分源变量的变量信息



值标签指明变量是二分的，值 0 和 1 分别表示 No 和 Yes。列表中的所有五个值都以相同方式进行编码，值 1（Yes 的代码）是多二分集的已计算的值。

类别

多类别集由多个变量组成，所有这些变量都以相同方式进行编码，通常具有许多可能的响应类别。例如，某个调查项目为“请列举最能描述您的种族血统的民族，最多三个”。可能有上百种回答，但为了进行编码，列表限制为 40 个最常见的民族，任何其他回答都归为“其他”类别。在数据文件中，三种选择成为三个变量，每个变量有 41 个类别（40 个已编码的民族和一个“其他”类别）。

在样本数据文件中，$ethmult 和 $mltcars 是多类别集。

类别标签源

对于多个二分，可以控制标记集合的方式。

- **变量标签**。使用已定义的变量标签（或不带已定义变量标签的变量的变量名）作为集类别标签。例如，如果集中的所有变量对于已计算的值（例如，Yes）具有相同的值标签（或未定义值标签），则应使用变量标签作为集类别标签。

- **已计算的值的标签。**使用已计算的值的已定义值标签作为集类别标签。只有在所有变量对于已计算的值都定义了值标签，且已计算的值的值标签对于每个变量都不相同时选择此选项。
- **使用变量标签作为集标签。**如果选择**已计算的值的标签**，则也可以使用集中具有已定义变量标签的第一个变量的变量标签作为集标签。如果集中的变量都未定义变量标签，则将集中第一个变量的名称作为集标签。

复制数据属性

通过复制数据属性向导可以使用外部 IBM® SPSS® Statistics 数据文件作为模板来定义活动数据集中的文件和变量属性。您也可以将活动数据集中的变量用作活动数据集中其他变量的模板。您可以：

- 将所选文件属性从外部数据文件或打开的数据集复制到活动数据集。文件属性包括文档、文件标签、多重响应集、变量集和加权。
- 将所选变量属性从外部数据文件或打开的数据集复制到活动数据集中匹配的变量。变量属性包括值标签、缺失值、测量级别、变量标签、打印和书写格式、对齐方式和列宽（在数据编辑器中）。
- 将所选变量属性从外部数据文件、打开的数据集或活动数据集中的一个变量复制到活动数据集中的多个变量。
- 根据外部数据文件或打开的数据集中的所选变量，在活动数据集中创建新变量。

在复制数据属性时，适用以下一般规则：

- 如果您使用外部数据文件作为源数据文件，则该文件必须为 SPSS Statistics 格式的数据文件。
- 如果使用活动数据集作为源数据文件，则其必须包含至少一个变量。不能使用完全空白的活动数据集作为源数据文件。
- 源数据集中的未定义（空）属性不会覆盖活动数据集中的已定义属性。
- 源变量中的变量属性将仅复制到类型匹配的目标变量 — 字符串（字母数值）或数值（包括数值、日期和货币）。

注意：“复制数据属性”替换了以前在“文件”菜单上提供的“应用数据字典”。

复制数据属性

▶ 从数据编辑器窗口的菜单中选择：

数据 > 复制数据属性...

图片 7-9
复制数据属性向导：第 1 步



▶ 选择具有您想要复制的文件和/或变量属性的数据文件。这可以是当前打开的数据集、外部 IBM® SPSS® Statistics 数据文件或活动数据集。

▶ 按照复制数据属性向导中的逐步操作说明进行操作。

选择源变量和目标变量

在这一步中，您可以指定包含想要复制的变量属性的源变量，以及将接收这些变量属性的目标变量。

图片 7-10
复制数据属性向导：第 2 步



将所选源数据集变量的属性应用于匹配的活动数据集变量。 将变量属性从一个或多个所选源变量复制到活动数据集中匹配的变量。如果变量名称和类型（字符串或数值）都相同，则变量是“匹配的”。对于字符串变量，定义的长度也必须相同。缺省情况下，在两个变量列表中仅显示匹配的变量。

- **如果尚不存在匹配变量，则在活动数据集中创建匹配变量。** 这将更新源列表以显示源数据文件中的所有变量。如果活动数据集中没有所选的源变量（基于变量名），则将在活动数据集中创建新变量，变量名和属性都出自源数据文件。

如果活动数据集不包含任何变量（空白的新数据集），则显示源数据文件中的所有变量，并基于所选源变量在活动数据集中自动创建新变量。

选择源列表中将复制属性的一个变量以及目标列表中将应用属性的一个或多个变量。 可以将源列表中的单个所选变量中的变量属性应用于活动数据集列表中的一个或多个所选变量。活动数据集列表中仅示与源列表中所选变量类型相同（数值或字符串）的变量。对于字符串变量，仅显示定义长度与源变量相同的字符串。如果活动数据集不包含变量，则此选项不可用。

注意：不能使用此选项在活动数据集中创建新变量。

仅应用数据集属性 — 未选定变量。 仅将文件属性（例如：文档、文件标签、权重）应用于活动数据集。将不应用任何变量属性。如果活动数据集同时也是源数据文件，则此选项不可用。

选择要复制的变量属性

您可以将选定的变量属性从源变量复制到目标变量。源变量中未定义的（空）属性不会覆盖目标变量中的已定义属性。

图片 7-11
复制数据属性向导：第 3 步



值标签。 值标签是与数据值相关联的描述性标签。当数据值用于代表非数值类别时（例如：代码 1 和 2 代表 Male 和 Female），常常使用值标签。您可以在目标变量中替换或合并值标签。

- **替换**删除目标变量的任何已定义的值标签，并用源变量中定义的值标签替换它们。
- **合并**将源变量中的已定义值标签与目标变量的任何现有已定义值标签进行合并。如果相同的值在源变量和目标变量中都有一个已定义值标签，则目标变量中的值标签保持不变。

定制属性。 用户定义的定制变量属性。

- **替换**删除目标变量的任何自定义属性，并用源变量中定义的属性替换它们。
- **合并**将源变量中定义的属性与目标变量的任何现有已定义属性进行合并。

缺失值。 缺失值是被标识为代表缺失数据的值（例如：98 代表 Do not know，99 代表 Not applicable）。通常，这些值也具有已定义的值标签，描述缺失值代码代表的含义。删除目标变量的所有现有的已定义缺失值，并用源变量中的已定义缺失值替换它们。

变量标签。 描述性变量标签可以包含空格和变量名称中所不允许的保留字符。如果正在将单个源变量中的变量属性复制到多个目标变量，则在选择此选项前可能需要三思。

测量级别。 测量级别可以是名义、有序或刻度。

角色。 某些对话框支持基于定义的角色预先选择分析变量的功能。

格式。 对于数值变量，此选项控制数值类型（如数值、日期或货币）、宽度（显示字符的总数，包括前导和拖尾字符以及小数指示符）和显示的小数位数。对于字符串变量，忽略此选项。

对齐。 此选项仅影响数据编辑器的“数据视图”中的对齐方式（左对齐、右对齐、居中）。

数据编辑器列宽度。 此选项仅影响数据编辑器的“数据视图”中的列宽。

复制数据集（文件）属性

可以将源数据文件中所选的全局数据集属性应用于活动数据集。（如果活动数据集是源数据文件，则此选项不可用。）

图片 7-12
复制数据属性向导：第 4 步



多重响应集。 将源数据文件中的多重响应集定义应用于活动数据集。

- 忽略在活动数据集中没有包含变量的多重响应集，除非将根据“复制数据属性向导”的第 2 步（选择源变量和目标变量）中的规范来创建这些变量。
- **替换**删除活动数据集中的所有多重响应集，并用源数据文件中的多重响应集替换它们。
- **合并**将源数据文件中的多重响应集添加到活动数据集中多重响应集的集合中。如果在两个文件中都存在相同名称的集合，则活动数据集中的现有集合保持不变。

变量集。 变量集用于控制在对话框中显示的变量列表。变量集是通过从“实用程序”菜单中选择“定义变量集”来定义的。

- 忽略源数据文件中包含活动数据集中没有的变量的集合，除非将根据复制数据属性向导的第 2 步（选择源变量和目标变量）中的规范来创建这些变量。
- **替换**删除活动数据集中任何现有变量集，并用源数据文件中的变量集替换它们。
- **合并**将源数据文件中的变量集添加到活动数据集中变量集的集合中。如果在两个文件中都存在相同名称的集合，则活动数据集中的现有集合保持不变。

文档。 通过 DOCUMENT 命令追加到数据文件中的注释。

- **替换**删除活动数据集中的任何现有文档，并用源数据文件中的文档替换它们。
- **合并**可用于合并源数据集和活动数据集中的文档。将在活动数据集中没有的源文件中的唯一文档添加到活动数据集。然后对所有文档按日期排序。

定制属性。 定制数据文件属性，通常使用命令语法中的 DATAFILE ATTRIBUTE 命令创建这些属性。

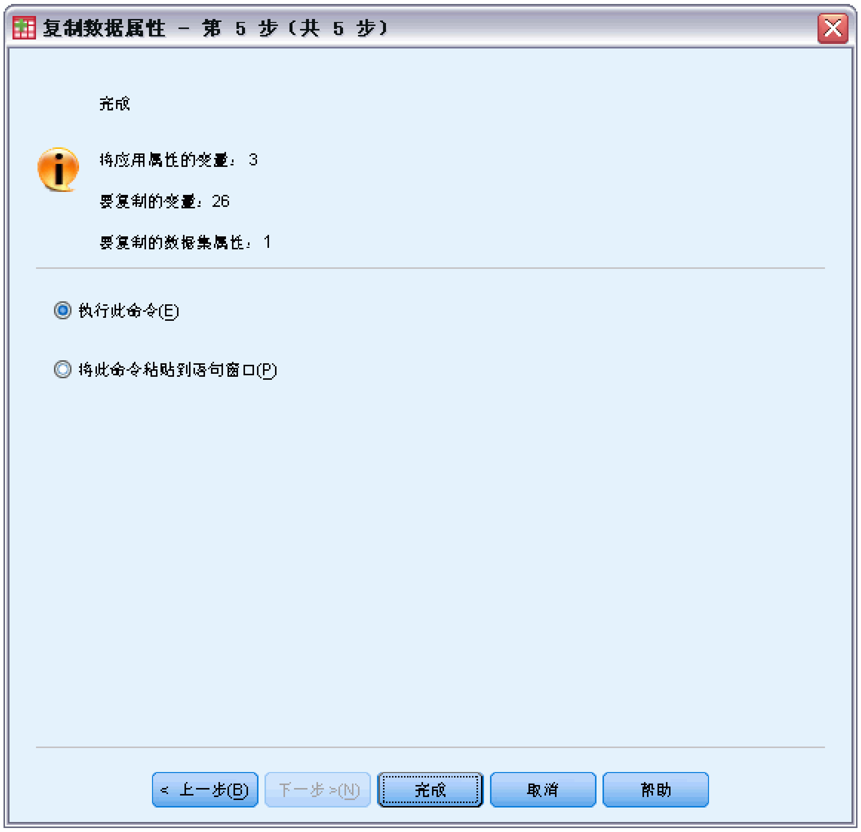
- **替换**删除活动数据集中的任何现有定制数据文件属性，并用源数据文件中的数据文件属性替换它们。
- **合并**可用于合并源数据集和活动数据集中的数据文件属性。将活动数据集中没有的源文件中的唯一属性名添加到活动数据集。如果在两个数据文件中都存在相同的属性名，则活动数据集中的已命名属性保持不变。

权重指定。 如果活动数据集中有匹配的变量，则按源数据文件中的当前权重变量对个案进行加权。这会覆盖活动数据集中当前生效的任何加权。

文件标签。 通过 FILE LABEL 命令向数据文件应用的描述性标签。

结果

图片 7-13
复制数据属性向导：第 5 步



复制数据属性向导中的最后一步提供如下信息：变量数目（将从源数据文件中复制这些变量的变量属性）、将创建的新变量的数目和将复制的数据集（文件）属性的数目。

也可以选择将所生成的命令语法粘贴到语法窗口中，并保存该语法以供以后使用。

标识重复个案

在数据中出现“重复”个案有多种原因，包括：

- 数据输入错误，意外地多次输入了同一个案。
- 多个个案具有相同的主标识值，但它们有不同的次标识值，就像居住在同一间屋子的多个家庭成员。
- 多个个案代表同一个案，但是对于除标识该个案的变量之外的其他变量有不同值，例如由同一个人或公司在不同时间购买的不同产品。

“标识重复个案”允许您自由定义重复，并在一定程度上控制对主个案和重复个案的自动确定。

标识和标记重复个案

▶ 从菜单中选择：
数据 > 标识重复个案...

▶ 选择标识匹配个案的一个或多个变量。

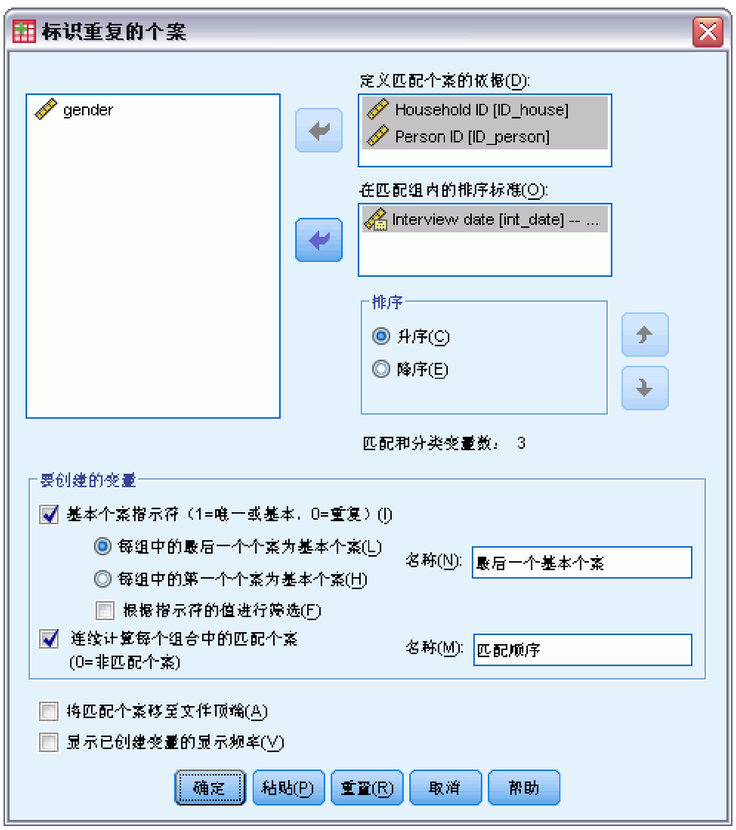
▶ 在“要创建的变量”组中选择一个或多个选项。

根据需要，您可以：

▶ 选择一个或多个变量对由选定的匹配个案变量定义的组内的个案排序。由这些变量定义的排序顺序决定了每一组中的“第一个”和“最后一个”个案。否则，将使用原文件顺序。

▶ 自动过滤重复个案使其不包含在报告、图表或统计量计算中。

图片 7-14
“标识重复的个案”对话框



定义匹配个案的依据。 如果个案的值匹配所有选定的变量，则它们视为是重复个案。如果您只想标识在所有方面 100% 匹配的个案，请选中所有变量。

在匹配组内的排序标准。 将按定义匹配个案的变量自动对个案排序。您可以选择将决定每一匹配组中个案的顺序的其他排序变量。

■ 对每个排序变量，您可按升序或降序排序。

- 如果您选择多个排序变量，将按列表中前一个变量类别中的每个变量对个案进行排序。例如，如果您选择日期作为第一排序变量，数量作为第二排序变量，则在每一日期中将按数量对个案排序。
- 使用列表右边的向上和向下箭头按钮更改变量的排序顺序。
- 排序顺序决定了每个匹配组内的“第一个”和“最后一个”个案，这些个案将决定可选的主指示变量的值。例如，如果您想要过滤每一匹配组中除最近个案之外的所有个案，您可以按日期变量的升序对组内个案排序，这将使得组中最近的日期排在最后。

基本个案指示符。 为所有唯一个案和每一匹配个案组中标识为主个案的个案创建值为 1 的变量，为每一组中非主要的重复个案创建值为 0 的变量。

- 由匹配组内的排序顺序确定，主个案可为每一匹配组中的最后一个或是第一个个案。如果未指定任何排序变量，原文件顺序将决定每一组内的个案顺序。
- 您可将指示变量作为**过滤变量**，在报告和分析中排除非主要的重复个案，而不从数据文件中删除这些个案。

每组中匹配个案的连续计数。 在每一匹配组中为个案创建序列值为 1 到 n 的变量。该序列基于每一组中当前个案的顺序，可以是原文件顺序，也可以是由任何指定的排序变量决定的顺序。

将匹配个案移至文件顶端。 对数据文件排序使所有匹配个案组位于数据文件顶部，这样方便在数据编辑器中查看匹配个案。

显示已创建变量的显示频率。 包含已创建变量的每一个值的计数的频率表。例如，对主指示变量，表中显示变量值为 0 的个案数（表示重复个案数）和变量值为 1 的个案数（表示唯一主个案数）。

缺失值。 对于数值变量，将系统缺失值视为和任何其他值一样—即将标识变量有系统缺失值的个案视为该变量具有匹配值的个案。对字符串变量，标识变量没有值的个案将作为该变量具有匹配值的个案。

可视离散化

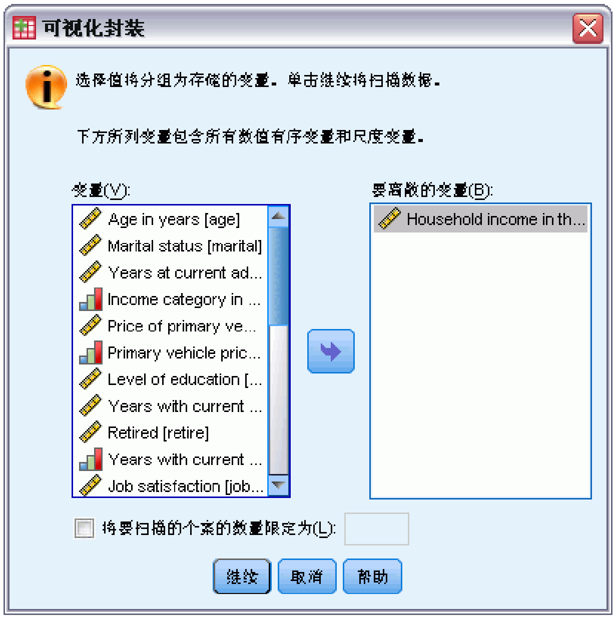
可视离散化设计的目的在于，帮助您在将现有变量的连续值进行分组的基础上，将新变量创建到数目有限的不同类别中。可以将可视离散化用于：

- 从连续刻度变量创建分类变量。例如，您可以使用刻度收入变量创建包含收入范围的新的分类变量。
- 将大量有序类别拼并到一小组类别中。例如，您可以将具有 9 个级别的等级标度拼并为分别代表低、中、高的三个类别。

在第一步中：

▶ 选择要为其创建新类别（离散化）变量的数值尺度和/或序数变量。

图片 7-15
用于选择要离散化的变量的初始对话框



您可以选择限制要扫描的个案数。对于含有大量个案的数据文件，限制扫描个案的数目可以节省时间，但您应尽可能避免限制此数目，因为这会影响可视离散化中的后续计算使用的值的分布。

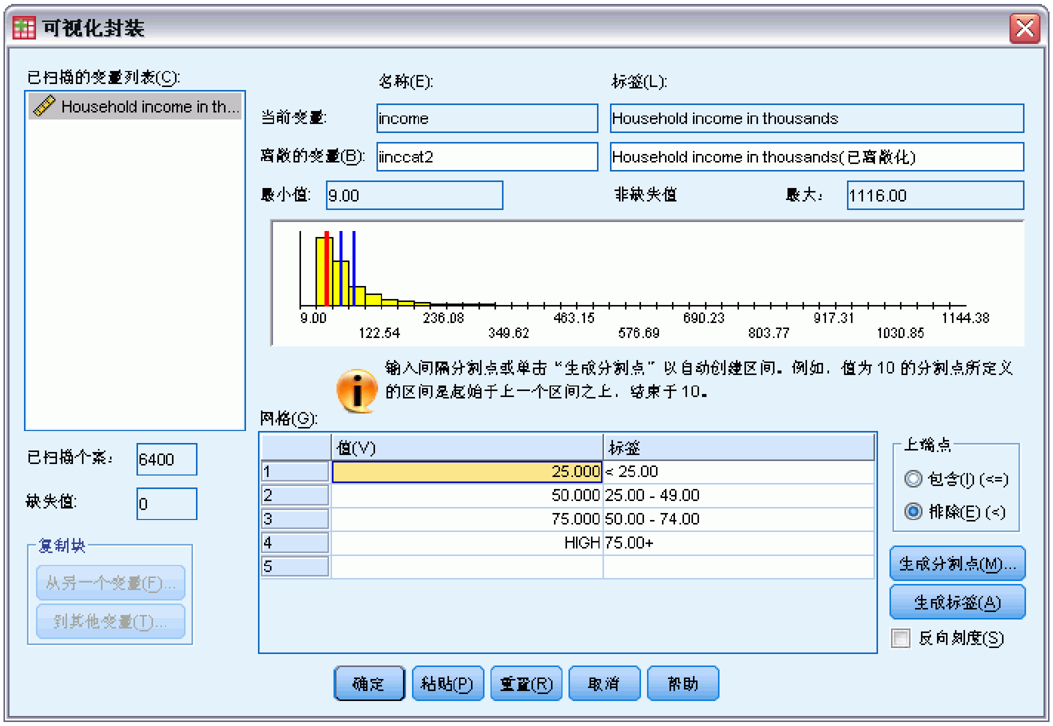
注意：字符串变量和标定数值变量不显示在源变量列表中。可视离散化需要按标度或序数级别度量的数值变量，因为它假定数据值代表某种逻辑顺序，该顺序可用于以有意义的方式对值分组。您可以在数据编辑器的变量视图中更改变量的定义测量级别。

离散化变量

- 从数据编辑器窗口的菜单中选择：
  转换 > 可视离散化...
- 选择要为其创建新类别（离散化）变量的数值尺度和/或序数变量。
- 在“已扫描的变量”列表中选择一个变量。
- 为新的离散化变量输入一个名称。变量名必须是唯一的，并且必须遵循变量命名规则。
- 为新变量定义离散化标准。
- 单击**确定**。

离散化变量

图片 7-16
可视离散化主对话框



可视离散化主对话框提供已扫描变量的以下信息：

已扫描的变量列表。显示您在初始对话框中选择的变量。通过单击列标题，您可以按测量级别（标度或有序）、变量标签或名称对列表排序。

已扫描的个案。指示已扫描个案的数目。将使用选定变量的所有不带用户缺失值或系统缺失值的已扫描个案，来生成在可视离散化的计算中使用的值的分布，包括在主对话框中显示的直方图，以及基于百分位或标准差单位的分割点。

缺失值。指示带有用户缺失值或系统缺失值的已扫描个案的数目。缺失值不包含在任何离散化类别中。

当前变量。当前选定变量的名称和变量标签（如果有），该变量将用作新的离散化变量的基础。

离散的变量。新的离散化变量的名称和可选的变量标签。

- **名称。**必须为新变量输入一个名称。变量名必须是唯一的，并且必须遵循变量命名规则。
- **标签。**您可以输入长度不超过 255 个字符的描述性变量标签。缺省的变量标签是标签末尾附有（离散化）的源变量的变量名或变量标签（如果有）。

最小值和最大值。当前选定变量的最小值和最大值，它基于已扫描的个案，并且不包括定义为用户缺失的值。

非缺失值。此直方图基于已扫描个案，显示当前选定变量的非缺失值的分布。

- 在为新变量定义了块之后，会在直方图上显示垂直线表示定义块的分割点。
- 您可以单击分割点线并拖到直方图上的其他位置，以更改块范围。
- 您还可以通过将分割点线拖到直方图外而移去块。

注意：（显示非缺失值的）直方图、最小值和最大值均基于已扫描的值。如果不在扫描中包含所有个案，就可能不能精确地反映真实的分布，尤其是对于数据文件已按选定变量排序的情况。如果不扫描任何个案，则不提供任何关于值的分布的信息。

网格。显示定义每个块的上端点的值，以及每个块的可选值标签。

- **值**。定义每个块的上端点的值。您可以输入值，或者使用**生成分割点**以基于选定的标准自动创建块。缺省情况下，会自动包含值为 HIGH 的分割点。此块将包含其他分割点以上的所有非缺失值。由最低分割点定义的块将包含所有低于或等于该值的非缺失值（或者只包含低于该值的非缺失值，这取决于您如何定义上端点）。
- **标签**。新的离散化变量的值的可选描述性标签。由于新变量的值只是从 1 到 n 的连续整数，所以描述值所代表的内容的标签非常有用。您可以输入标签，或者使用**生成标签**以自动创建值标签。

从网格中删除块

- ▶ 右键单击块的值或标签单元格。
- ▶ 从弹出的上下文菜单中，选择**删除行**。

注意：如果删除 HIGH 块，那么值高于最后一个指定分割点值的所有个案对于新变量都被分配系统缺失值。

删除所有标签或删除所有定义的块

- ▶ 右键单击网格中的任意位置。
- ▶ 从弹出的上下文菜单中，选择**删除所有标签**或**删除所有分割点**。

上端点。控制在网格的值列输入的上端点值的处理。

- **包含 (<=)**。具有在值单元格中指定的值的个案包含在离散化类别中。例如，如果您指定值 25、50 和 75，则值正好为 25 的个案将进入第一个块，因为它将包含值小于等于 25 的所有个案。
- **排除 (<)**。具有在值单元格中指定的值的个案不包含在离散化类别中。而是包含在下一个块中。例如，如果您指定值 25、50 和 75，则值正好为 25 的个案将进入第二个块而非第一个块，因为第一个块将仅包含值小于 25 的个案。

生成分割点。自动为等宽度间隔、具有相等个案数的区间或者基于标准差的区间生成离散化类别。如果未扫描任何个案，则此项不可用。

生成标签。基于网格中的值和指定的上端点的处理（包含或排除），为新的离散化变量的连续整数值生成描述性标签。

反向刻度。缺省情况下，新的离散化变量的值是从 1 到 n 的升序整数。反向刻度会使得这些值成为从 n 到 1 的降序整数。

复制块。您可以将离散化规范从另一个变量复制到当前选定变量，或者从选定变量复制到多个其他变量。

自动生成离散化类别

“生成分割点”对话框允许您基于选定的标准自动生成离散化类别。

使用“生成分割点”对话框

- ▶ 在“已扫描的变量列表”中选择（单击）一个变量。
- ▶ 单击**生成分割点**。
- ▶ 选择用于生成将定义离散化类别的分割点的标准。
- ▶ 单击**应用**。

图片 7-17
“生成分割点”对话框



注意：如果未扫描到任何个案，则“生成分割点”对话框不可用。

等宽度间隔。基于以下三个标准中的任意两个，生成等宽（例如 1 – 10、11 – 20 和 21 – 30）的离散化类别：

- **第一个分割点的位置。**定义最下面的离散化类别的上端点的值（例如，值 10 表示包含所有不超过 10 的值的范围）。
- **分割点数量。**离散化类别数是分割点数量加一。例如，9 个分割点会生成 10 个离散化类别。
- **宽度。**每个区间的宽度。例如，值 10 会将年龄 离散化为长度为 10 年的区间。

基于已扫描个案的等百分位。基于以下标准之一，生成一些离散化类别，使得每个块中的个案数相等（对于百分位，使用 Aempirical 算法）：

- **分割点数量。**离散化类别数是分割点数量加一。例如，三个分割点会生成四个百分位块（四分位数），每个块包含 25% 的个案。
- **宽度(%)。**每个区间的宽度，表示为个案总数的百分比。例如，值 33.3 将生成三个离散化类别（两个分割点），每个类别包含 33.3% 的个案。

如果源变量包含的不同值相对较少，或者有大量个案具有相同的值，则获取的块数可能少于请求的块数。如果在分割点处有多个相同的值，则它们都将转到相同的区间；因此实际百分比并不总是完全相同的。

基于已扫描个案的平均和选定标准差处的分割点。基于变量分布的均值和标准差的值生成离散化类别。

- 如果不选择任何标准差区间，则将使用均值作为分割点来划分块，从而创建两个离散化类别。
- 您可以基于一倍、两倍和/或三倍标准差选择标准差区间的任意组合。例如，选择所有三个标准差将生成八个离散化类别 — 每个标准差区间内两个块（三个区间共六个块），均值上下超过三倍标准差的个案两个块。

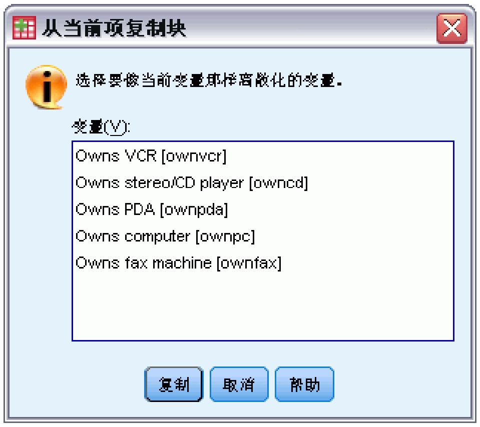
在正态分布中，68% 的个案落在均值的一倍标准差范围内；95% 的个案落在均值的两倍标准差范围内；99% 的个案落在均值的三倍标准差范围内。基于标准差创建离散化类别可能导致一些定义的块落在实际数据范围以外，甚至落在可能的数据值范围以外（例如，负的薪金范围）。

注意：百分位和标准差的计算是基于已扫描个案的。如果您限制已扫描个案的数目，则生成的块可能不包含一部分您本想在块中包含的个案，尤其是对于数据文件按源变量排序的情况。例如，如果您将扫描仅限于具有 1000 个个案的数据文件的前 100 个个案，且该数据文件按响应者年龄升序排序，那么，您会发现前三个块每个仅包含约 3.3% 的个案，而最后一个块包含 90% 的个案，而非四个百分位年龄块的每一个均包含 25% 的个案。

复制离散化类别

为一个或多个变量创建离散化类别时，您可以将离散化规范从另一个变量复制到当前选定变量，或者从选定变量复制到多个其他变量。

图片 7-18
将块复制到当前变量或者从当前变量复制块



复制离散化规范

- 为至少一个变量定义离散化类别，请勿 单击**确定**或**粘贴**。
- 在您已为其定义了离散化类别的“已扫描的变量列表”中，选择（单击）一个变量。
- 单击**到其他变量**。
- 选择您想要为其以相同的离散化类别创建新变量的变量。
- 单击**复制**。

或

- 在您想要向其复制定义的离散化类别的“已扫描的变量列表”中，选择（单击）一个变量。
- 单击**从另一个变量**。
- 选择您想要复制的具有定义的离散化类别的变量。
- 单击**复制**。

如果对于您从中复制离散化规范的变量，有指定的值标签，则也复制这些指定的值标签。

注意：在“可视离散化”主对话框中单击**确定**新建离散化变量（或者以任何其他方式关闭该对话框）后，就不能使用“可视离散化”将这些离散化类别复制到其他变量。

可视离散化中的用户缺失值

源变量中定义为用户缺失的值（标识为缺失数据代码的值）不包含在新变量的离散化类别中。源变量的用户缺失值复制为新变量的用户缺失值，同时还复制缺失值代码的任何已定义值标签。

如果某个缺失值代码与新变量的其中一个离散化类别值相冲突，则通过为最高离散化类别值加上 100，将新变量的缺失值代码重新编码为不冲突的值。例如，如果值 1 定义为源变量的用户缺失值，且新变量将有六个离散化类别，则对于源变量值为 1 的任何个

案，其新变量的值将为 106，且 106 将定义为用户缺失值。如果源变量的用户缺失值已有定义的值标签，则将保留该标签作为新变量的重新编码值的值标签。

注意：如果源变量有定义的用户缺失值范围，其形式为 LO-n，其中 n 是正数，那么新变量相应的用户缺失值将是负数。

章

8

数据转换

在理想情况下，原始数据非常适用于要执行分析的类型，并且，变量间的任何关系都是合适的线性或切合的正交关系。不幸的是，这种情况非常少。初步分析可能会暴露出编码方案不方便或编码错误，或者可能需要数据转换以揭示变量间的真实关系。

您可以执行从简单任务（比如拼并类别以进行分析）到更高级任务（比如基于复杂方程和条件语句创建新的变量）的数据转换。

计算变量

使用“计算”对话框可以基于其他变量的数值转换来计算变量的值。

- 您可以计算数值或字符串（字母数值）变量的值。
- 您可以创建新的变量或替换现有变量的值。对于新的变量，您也可以指定变量类型和标签。
- 您可以基于逻辑条件有选择的计算数据子集的值。
- 您可以使用多种内置函数，包括算术函数、统计函数、分布函数和字符串函数。

图片 8-1
“计算变量”对话框



计算变量

▶ 从菜单中选择：

转换 > 计算变量...

▶ 输入一个目标变量的名称。它可以是一个现有变量或是一个要添加到活动数据集中的新变量。

▶ 要构建一个表达式，可以将成分粘贴到“表达式”字段中或是在“表达式”字段中直接输入。

- 通过从“函数组”列表中选择组，然后双击“函数和特殊变量”列表中的函数或变量（或选择函数或变量，然后单击“函数组”列表相邻的箭头），可以粘贴函数或常用的系统变量。填充问号指示的任何参数（仅适用于函数）。标记为**所有**的函数组提供所有可用函数和系统变量的列表。对话框的保留区域中显示对当前所选函数或变量的简要描述。
- 字符串常数必须包含在引号或撇号中。
- 如果值包含小数，则必须使用句号（.）作为小数指示符。
- 对于新的字符串变量，必须选择**类型和标签**以指定数据类型。

计算变量：If 个案

“If 个案”对话框允许您使用条件表达式对选定的个案子集应用数据转换。条件表达式对每个个案都返回一个值：true、false 或 missing。

图片 8-2
“计算变量：If 个案”对话框



- 如果条件表达式的结果为 true，则所选子集中将包含该个案。
- 如果条件表达式结果为 false 或 missing，则所选子集中不包含该个案。
- 大多数条件表达式使用计算器面板上六个关系运算符（<、>、<=、>=、= 和 ~=）中的一个或多个。
- 条件表达式可以包含变量名称、常数、算术运算符、数值（和其他）函数、逻辑变量以及关系运算符。

计算变量：类型和标签

默认情况下，新的计算变量是数值变量。要计算新的字符串变量，必须指定数据类型和宽度。

标签。 描述性变量标签最多可以包含 255 个字节。您可以输入标签或使用计算表达式的前 110 个字符作为标签。

类型。计算变量可以是数值或字符串（字母数值）。字符串变量不能用于计算。

图片 8-3
“类型和标签”对话框



函数

支持很多类型的函数，包括：

- 算术函数
- 统计函数
- 字符串函数
- 日期和时间函数
- 分布函数
- 随机变量函数
- 缺失值函数
- 评分函数

若要了解上述各函数的更多信息和详细说明，请在帮助系统的“索引”选项卡中键入函数。

函数中的缺失值

函数和简单算术表达式以不同的方式处理缺失值。在以下表达式中：

```
(var1+var2+var3)/3
```

如果一个个案里三个变量中的任何一个有缺失值，结果将缺失。

在以下表达式中：

```
MEAN(var1, var2, var3)
```

只有当个案中三个变量都有缺失值的时候，结果才将缺失。

对于统计函数，您可以指定必须包含非缺失值的参数的最小数目。为此，要在函数名称之后输入一个句点和最小数目，例如：

```
MEAN.2(var1, var2, var3)
```

随机数字生成器

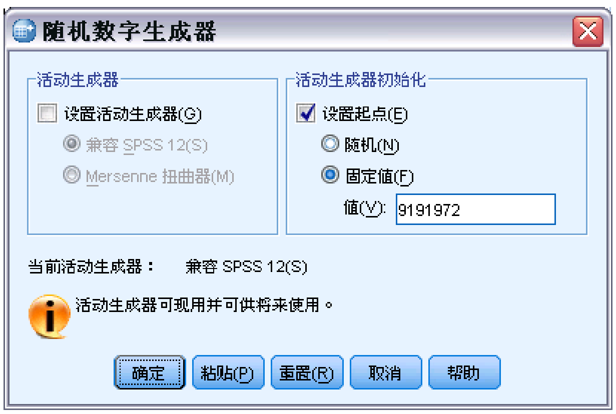
使用“随机数生成器”对话框可以选择随机数生成器并设置起始序列值，以便重新生成随机数的序列。

活动生成器。有两种不同的随机数字生成器可供使用：

- **兼容版本 12.** 版本 12 和早期版本中所使用的随机数字生成器。如果您要重新生成在早期版本中根据指定的种子值生成的随机结果，可使用此随机数字生成器。
- **Mersenne 扭曲器.** 较新的随机数生成器，可实现更可靠的模拟。如果从版本 12 或早期版本再现随机结果不是问题，那么可使用此随机数生成器。

活动生成器初始化。每次生成随机数以用于转换（例如随机分布函数）、随机抽样或是个案加权时，随机数种子都随之更改。要复制随机数的序列，请先设置初始化起始点值，然后再进行使用这些随机数的每个分析。值必须为正整数。

图片 8-4
“随机数生成器”对话框



一些过程具有内部随机数生成器，包括：

选择随机数生成器和/或设置初始化值：

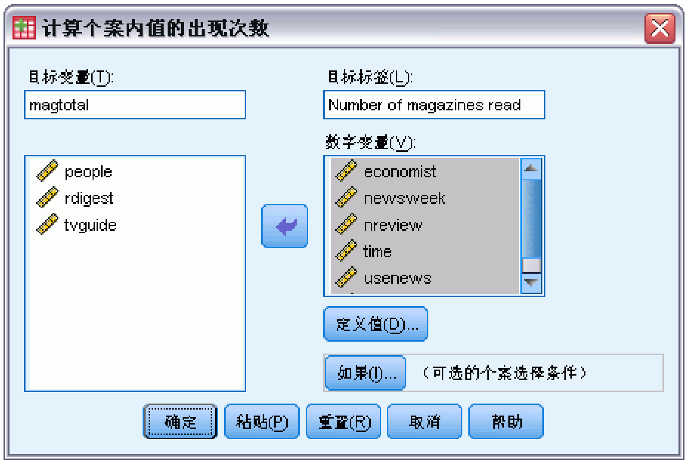
▶ 从菜单中选择：

转换 > 随机数字生成器

计算个案内值的出现次数

该对话框将创建一个变量，该变量统计每个个案的变量列表中相同值的出现次数。例如，某调查可能包含一个杂志列表，并使用是/否复选框来表示每个响应者阅读哪些杂志。您可以计算响应者回答是的数目以创建包含他所阅读的杂志总数的一个新变量。

图片 8-5
“计算个案内值的出现次数”对话框



计算个案内值的出现次数

▶ 从菜单中选择：
转换 > 对个案内的值计数...

▶ 输入一个目标变量名称。

▶ 选择同类型（数值或字符串）的两个或更多变量。

▶ 单击**定义值**并指定应计数的一个或多个值。

或者，可以定义要为其计算值出现次数的个案子集。

计算个案内的值：要计数的值

每当选定变量之一与此处的“要计数的值”列表中指定的值匹配时，目标变量（主对话框中）的值就增加 1。如果一个个案匹配任何变量的多个指定值，则目标变量就为该变量递增多次。

指定的值可包括个别值、缺失值或系统缺失值和范围。范围包括其端点和范围内的任何用户缺失值。

图片 8-6
“要统计的值”对话框



统计出现次数：If 个案

通过“If 个案”对话框可以使用条件表达式统计所选个案子集的值的出现次数。条件表达式对每个个案都返回一个值：true、false 或 missing。

图片 8-7
“统计出现次数：If 个案”对话框



有关使用“If 个案”对话框的一般注意事项，请参见计算变量：If 个案第 121 页码。

转换值

转换值创建包含来自前面或后面个案的现有变量值的新变量。

名称。 新变量的名称。这必须是活动数据集中尚未存在的名称。

从早期个案（延迟）获取值。 从活动数据集中前一个个案获取值。例如，当缺省个案数是 1 时，新变量的每个个案有来自前一个个案的原始变量的值。

从以下个案（提前）获取值。 从活动数据集中后一个个案获取值。例如，当缺省个案数是 1 时，新变量的每个个案有来自下一个个案的原始变量的值。

要转换的个案数。 从前面或后面第 n 个个案获取值，其中 n 是指定的值。值必须是一个非负整数。

- 如果分割文件处理正在进行，转换的范围限于每个分割组。无法从前面或后面分割组中的个案获取转换值。
- 忽略过滤状态。

- 结果变量的值被设置为数据集或分割组中最开始或最后 n 个个案的系统缺失，其中 n 是为**要转换的个案数**指定的值。例如，使用值为 1 的延迟方法会将结果变量设置为数据集中第一个个案（或每个分割组中第一个个案）的系统缺失。
- 保留用户缺失值。
- 来自原始变量的字典信息（包括定义值标签和用户缺失值分配）被应用于新变量。（注意：不包括定制变量属性。）
- 对于描述创建变量的转换操作的新变量，自动生成变量标签。

创建具有转换值的新变量

▶ 从菜单中选择：
转换 > 转换值

▶ 选择变量，用作新变量的值源。

▶ 输入新变量的名称。

▶ 选择转换方法（延迟或提前）和要转换的个案数。

▶ 单击**更改**。

▶ 为您想创建的每个新变量重复该操作。

对值重新编码

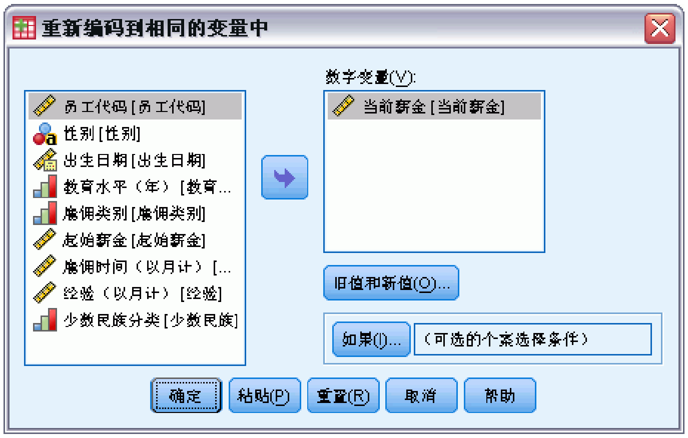
您可通过对值重新编码来修改数据值。这对拼并类别或组合类别特别有用。您可对现有变量中的值重新编码，或基于现有变量的重新编码后的值创建新的变量。

重新编码到相同的变量中

使用“重新编码到相同的变量中”对话框可以重新指派现有变量的值或将现有值的范围拼并为新值。例如，可将薪金拼并到薪金范围类别中。

您可以对数值变量和字符串变量重新编码。如果选择多个变量，则它们必须为相同类型。不可将数值变量和字符串变量放在一起重新编码。

图片 8-8
“重新编码到相同的变量中”对话框



对变量的值重新编码

▶ 从菜单中选择：
转换 > 重新编码为相同变量...

▶ 选择要重新编码的变量。如果选择多个变量，则它们必须为相同类型（数值或字符串）。

▶ 单击**旧值和新值**并指定如何对值重新编码。

或者，您可定义要重新编码的个案子集。用于进行此操作的“If 个案”对话框与为“统计出现次数”描述的对话框相同。

重新编码为相同变量：旧值和新值

您可在该对话框中定义要重新编码的值。所有指定的值必须与主对话框中选择的变量的数据类型（数值或字符串）相同。

旧值。要重新编码的值。您可对单个值、值范围和缺失值重新编码。不可为字符串变量选择系统缺失值和范围，因为这两个概念都不适用于字符串变量。范围包括其端点和范围内的任何用户缺失值。

- **值**. 要重新编码为新值的单个旧值。该值必须具有与所重新编码的变量相同的数据类型（数值或字符串）。
- **系统缺失**. 在以下各种情况下由程序指定的值：未根据您已指定的格式类型定义数据中的值，数值字段为空，或者未定义从转换命令生成的值。数值型的系统缺失值显示为句号。字符串变量不能具有系统缺失值，因为任何字符在字符串变量中均是合法的。
- **系统或用户缺失**. 满足以下条件的观察值：其值定义为用户缺失值，或者未知且被指定了系统缺失值（以句点 . 表示）。
- **范围**. 值的包含范围。不适用于字符串变量。包含范围内的所有用户缺失值。
- **所有其他值**. 未包含在“旧-新”列表上其中一个规范中的所有剩余值。这在“旧-新”列表上显示为“ELSE”。

新值。由每个旧值或值范围重新编码获得的单个值。您可输入一个值或指定系统缺失值。

- **值.** 一个或多个旧值将要重新编码为的值。该值必须具有与旧值相同的数据类型（数值或字符串）。
- **系统缺失.** 将指定的旧值重新编码为系统缺失值。系统缺失值不用在计算中，并且许多过程中都排除带系统缺失值的个案。不适用于字符串变量。

旧－>>新。将用于对变量重新编码的指定值列表。可以在列表中添加、更改和删除指定值。列表将基于指定的旧值按下列顺序自动排序：单值、缺失值、范围和其他值。如果更改列表中的重新编码指定，则过程自动对列表重新排序（如果必要）以保持此顺序。

图片 8-9
“旧值和新值”对话框

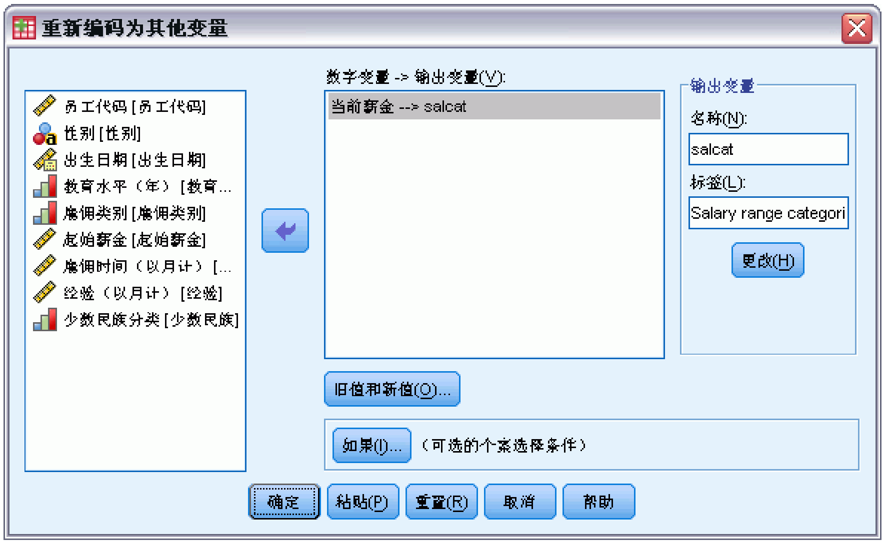


重新编码为其他变量

使用“重新编码为其他变量”对话框可以重新指派现有变量的值或将现有值的范围拼并为新变量的新值。例如，您可将薪金拼并到包含薪金范围类别的新变量中。

- 您可以对数值变量和字符串变量重新编码。
- 您可以将数值变量重新编码为字符串变量，反之亦然。
- 如果选择多个变量，则它们必须为相同类型。不可将数值变量和字符串变量放在一起重新编码。

图片 8-10
“重新编码为其他变量”对话框



将变量的值重新编码为新变量

▶ 从菜单中选择：
转换 > 重新编码为不同变量...

▶ 选择要重新编码的变量。如果选择多个变量，则它们必须为相同类型（数值或字符串）。

▶ 为每一新变量键入输出（新）变量名称并单击**更改**。

▶ 单击**旧值和新值**并指定如何对值重新编码。

或者，您可定义要重新编码的个案子集。用于进行此操作的“If 个案”对话框与为“统计出现次数”描述的对话框相同。

重新编码为不同变量：旧值和新值

您可在该对话框中定义要重新编码的值。

旧值。要重新编码的值。您可对单个值、值范围和缺失值重新编码。不可为字符串变量选择系统缺失值和范围，因为这两个概念都不适用于字符串变量。旧值必须与原变量是相同的数据类型（数值或字符串）。范围包括其端点和范围内的任何用户缺失值。

- **值**. 要重新编码为新值的单个旧值。该值必须具有与所重新编码的变量相同的数据类型（数值或字符串）。
- **系统缺失**. 在以下各种情况下由程序指定的值：未根据您已指定的格式类型定义数据中的值，数值字段为空，或者未定义从转换命令生成的值。数值型的系统缺失值显示为句号。字符串变量不能具有系统缺失值，因为任何字符在字符串变量中均是合法的。
- **系统或用户缺失**. 满足以下条件的观察值：其值定义为用户缺失值，或者未知且被指定了系统缺失值（以句点 . 表示）。

- **范围.** 值的包含范围。不适用于字符串变量。包含范围内的所有用户缺失值。
- **所有其他值.** 未包含在“旧-新”列表上其中一个规范中的所有剩余值。这在“旧-新”列表上显示为“ELSE”。

新值。由每个旧值或值范围重新编码获得的单个值。新值可为数值或字符串。

- **值.** 一个或多个旧值将要重新编码为的值。该值必须具有与旧值相同的数据类型（数值或字符串）。
- **系统缺失.** 将指定的旧值重新编码为系统缺失值。系统缺失值不用在计算中，并且许多过程中都排除带系统缺失值的个案。不适用于字符串变量。
- **复制旧值.** 可保留旧值。如果某些值不需要重新编码，则使用此项以包含旧值。任何未指定的旧值不包含在新变量中，具有那些值的个案将分配新变量的系统缺失值。

输出变量为字符串. 将重新编码的新变量定义为字符串（字母数字）变量。旧变量可以是数值或字符串。

将数值字符串移动为数值. 将包含数字的字符串值转换为数值。包含除数字和可选的符号（+ 或 -）以外的任何字符的字符串会分配系统缺失值。

旧 - >>新。将用于对变量重新编码的指定值列表。可以在列表中添加、更改和删除指定值。列表将基于指定的旧值按下列顺序自动排序：单值、缺失值、范围和其他值。如果更改列表中的重新编码指定，则过程自动对列表重新排序（如果必要）以保持此顺序。

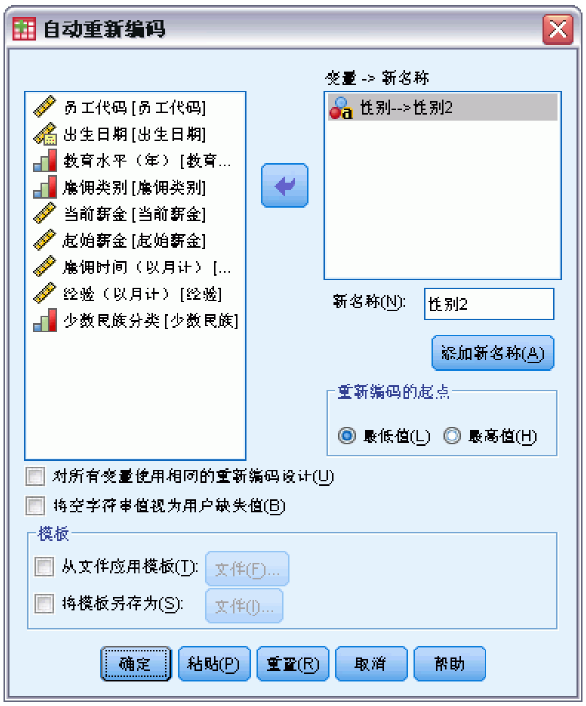
图片 8-11
“旧值和新值”对话框



自动重新编码

使用“自动重新编码”对话框可以将字符串值和数值转换为连续整数。当类别代码不连续时，对许多过程来说，生成的空单元将降低性能并增加内存要求。此外，某些过程不能使用字符串变量，某些过程要求因子水平为连续的整数值。

图片 8-12
“自动重新编码”对话框



- “自动重新编码”创建的新变量保留了旧变量中任何已定义的变量标签和值标签。对没有已定义值标签的任何值，将使用原值作为重新编码后的值的标签。一个表显示了旧值、新值以及值标签。
- 字符串值将按字母顺序重新编码，其中大写字母将排在相应的小写字母之前。
- 缺失值被重新编码为高于任何非缺失值的缺失值，并保留它们的原有顺序。例如，如果原变量有 10 个非缺失值，最低的缺失值将被重新编码为 11，值 11 将作为新变量的缺失值。

对所有变量使用相同的重新编码设计。使用此选项可以将一个自动重新编码设计应用于所有选定变量，从而为所有新变量生成一致的编码设计。

如果选择此选项，则下列规则和限制适用：

- 所有变量必须为相同类型（数值或字符串）。
- 所有选定变量的观察值用于创建要重新编码为连续整数的值的排序顺序。
- 新变量的用户缺失值基于具有已定义用户缺失值的列表中的第一个变量。将来自其他原变量的所有其他值（系统缺失值除外）都视为有效。

将空字符串值视为用户缺失值。对于字符串变量，不将空白值或空值视为**系统缺失值**。此选项会将空字符串自动重新编码为高于最高非缺失值的**用户缺失**值。

模板

可以将自动重新编码设计保存在模板文件中，然后将其应用于其他变量和其他数据文件。例如，每月可能有大量的字母数值产品代码要自动重新编码为整数，但在某些月份中添加了新产品代码，这会更改原始的自动重新编码设计。如果将原始设计保存在模板中，然后将其应用于包含一组新代码的新数据，则数据中遇到的任何新代码都会被自动重新编码为高于模板中最后一个值的值，从而保留了原始产品代码的原始自动重新编码设计。

将模板另存为。将所选变量的自动重新编码设计保存在外部模板文件中。

- 该模板包含将原始非缺失值映射到重新编码的值的信息。
- 模板中仅保存非缺失值的信息。不保留用户缺失值的信息。
- 如果选择了多个要进行重新编码的变量，但尚未选择对所有变量使用相同的自动重新编码设计，或者不应用现有模板作为自动重新编码的一部分，则模板将基于列表中的第一个变量。
- 如果选择了多个要进行重新编码的变量，还选择了**对所有变量使用相同的重新编码设计**和/或选择了**应用模板**，则模板将包含用于所有变量的组合自动重新编码设计。

从文件应用模板。将以前保存的自动重新编码模板应用于所选要进行重新编码的变量，从而将变量中的任何其他值附加到设计的末尾，并保留所保存设计中存储的原始值和自动重新编码的值的关系。

- 所选要进行重新编码的所有变量必须是相同类型（数值或字符串），并且该类型必须与模板中定义的类型匹配。
- 模板不包含有关用户缺失值的任何信息。目标变量的用户缺失值基于具有已定义用户缺失值的原变量列表中的第一个变量。将来自其他原变量的所有其他值（系统缺失值除外）都视为有效。
- 首先应用来自模板的值映射。将所有剩余的值重新编码为高于模板中最后一个值的值，同时将用户缺失值（基于具有已定义用户缺失值的列表中的第一个变量）重新编码为高于最后一个有效值的值。
- 如果选择了多个要进行自动重新编码的变量，则首先应用模板，然后对所选变量中的所有其他值进行通用的组合式自动重新编码，从而生成一个可用于所有所选变量的通用自动重新编码设计。

将字符串值或数字值重新编码为连续整数

▶ 从菜单中选择：
转换 > 自动重新编码...

▶ 选择一个或多个要重新编码的变量。

▶ 对每个选定的变量，为新变量输入一个名称并单击**新名称**。

个案排秩

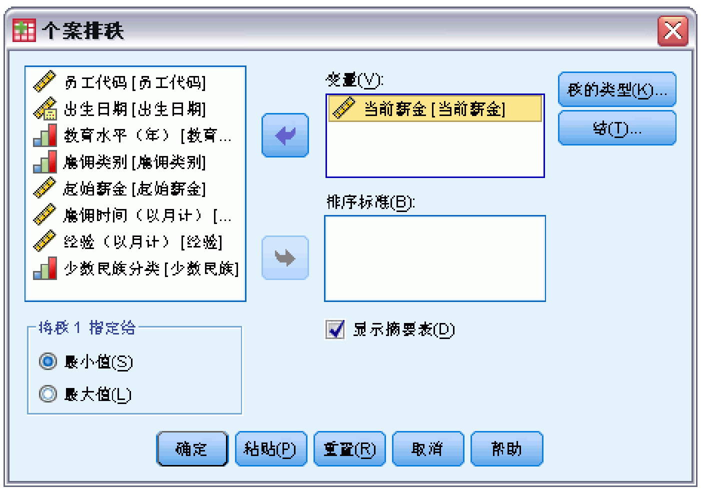
使用“个案排秩”对话框可以为数值变量创建包含秩、常规得分和 Savage 得分以及百分位值的新变量。

将基于原变量名称和选定的测量自动生成新变量名称和描述性变量标签。一个摘要表将列出原变量、新变量和变量标签。（注意：自动生成的新变量名称的最大长度为 8 个字节。）

根据需要，您可以：

- 按升序或降序对个案排秩。
- 通过在“依据”列表中选择一个或多个分组变量而将排秩组织为子组。将计算每个组中的秩。组通过分组变量的值组合定义。例如，如果您选择 gender 和 minority 作为分组变量，则将为 gender 和 minority 的每一组合计算秩。

图片 8-13
“个案排秩”对话框



对个案排秩

▶ 从菜单中选择：

转换 > 个案排秩...

▶ 选择一个或多个要排秩的变量。您可以只对数值变量排秩。

或者，您可按升序或降序对个案排秩并将秩组织为子组。

个案排秩：类型

您可以选择多种排秩方法。将为每种方法创建单独的排秩变量。排秩方法包含简单秩、Savage 得分、分数秩和百分位数。您也可以基于比例估计和正态得分创建排秩。

秩. 简单秩。新变量的值等于它的秩。

Savage 得分. 新变量包含基于指数分布的 Savage 得分。

分数秩. 新变量的值等于秩除以非缺失个案的权重总和。

百分比分数秩. 每个秩除以带有有效值的个案数，再乘以 100。

个案权重总和. 新变量的值等于个案权重的合计。对于同一组中的所有个案，该新变量是一个常数。

Ntile. 基于百分位组的秩，每一组包含的个案数大致相同。例如，4 个 Ntile 会将秩 1 指定给第 25 个百分位以下的个案，将秩 2 指定给第 25 个与第 50 个百分位之间的个案，将秩 3 指定给第 50 个与第 75 个百分位之间的个案，将秩 4 指定给第 75 个百分位以上的个案。

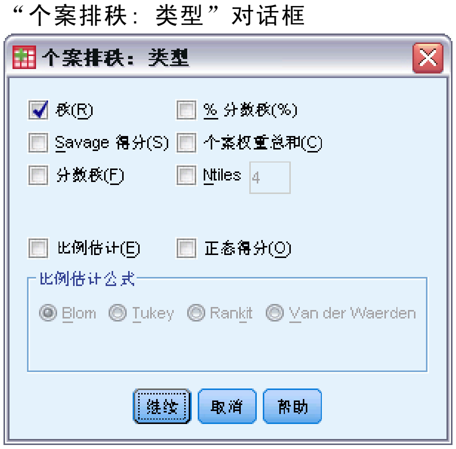
比例估计. 估计与特定秩对应的分布的累积比例。

正态得分. 对应于估计的累积比例的 z 得分。

比例估计公式。 对于比例估计和正态得分，您可以选择比例估计公式：Blom、Tukey、Rankit 或 Van der Waerden。

- **Blom**. 基于使用公式 (r-3/8) / (w+1/4) 的比例估计创建新的秩变量，其中 w 是个案权重的总和，r 是秩。
- **Tukey**. 使用公式 (r-1/3) / (w+1/3)，其中 r 为秩，w 为个案权重的总和。
- **Rankit**. 使用公式 (r-1/2) / w，其中 w 是观察次数，r 是秩，范围是从 1 到 w。
- **Van der Waerden**. Van der Waerden 转换，由公式 r/(w+1) 定义，其中 w 是个案权重的合计，r 是秩，范围从 1 到 w。

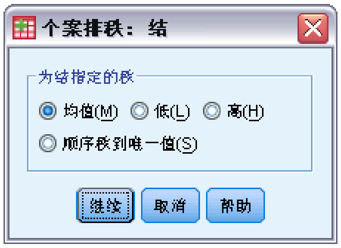
图片 8-14
“个案排秩：类型”对话框



个案排秩：结

该对话框控制为原变量上具有相同值的个案指定秩的方法。

图片 8-15
“个案排秩：同数”对话框



下表显示了不同方法是如何将秩指定给同数的值的：

值	均值	低	高	序列
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

日期和时间向导

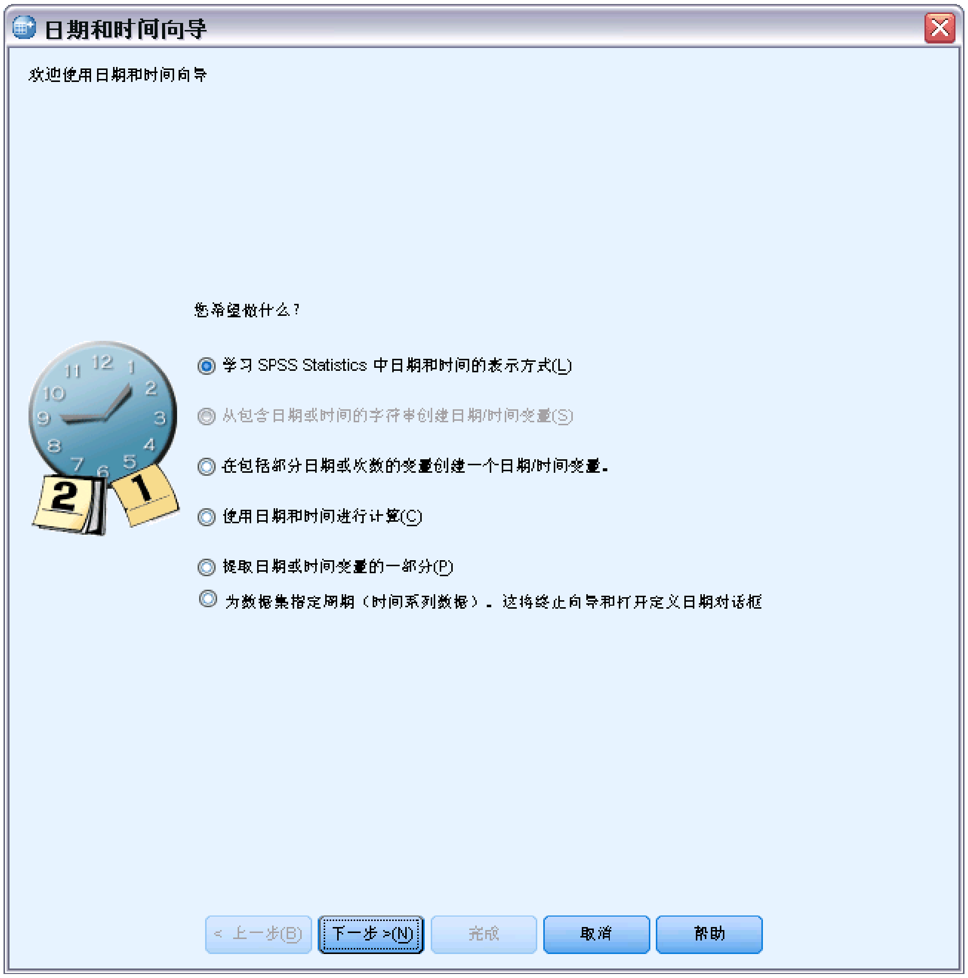
日期和时间向导简化了与日期和时间变量关联的许多常见任务。

使用日期和时间向导

▶ 从菜单中选择：
转换 > 日期和时间向导...

▶ 选择要完成的任务，并按照操作步骤定义该任务。

图片 8-16
日期和时间向导简介屏幕



- **学习日期和时间的表示方式。** 此选择将打开一个屏幕，其中提供 IBM® SPSS® Statistics 中日期/时间变量的简要概述。通过单击**帮助**按钮后，此屏幕还可提供指向更详细信息的链接。
- **从包含日期或时间的字符串中创建日期/时间变量。** 使用此选项从可用字符串变量中创建日期/时间变量。例如，某个字符串变量以 mm/dd/yyyy 的形式表示日期，您想要据此创建日期/时间变量。
- **在包括部分日期或次数的变量中创建一个日期/时间变量。** 使用此选项可以从一组现有变量中构建日期/时间变量。例如，一个变量表示月（作为整数），第二个变量表示一月中的一天，第三个变量表示年。可以将这三个变量组合成单个日期/时间变量。
- **使用日期和时间进行计算。** 使用此选项在日期/时间变量中加上或减去值。例如，可以通过从表示某过程的结束时间的变量中减去表示该过程的开始时间的另一变量，来计算该过程的期间。

- **提取日期或时间变量的一部分。** 使用此选项可以提取日期/时间变量的组成部分，如从一个日期/时间变量（形式为 mm/dd/yyyy）中提取一月中的一天。
- **为数据集指定周期。** 此选项会打开“定义日期”对话框，用于创建由一组连续日期组成的日期/时间变量。此功能通常用于将日期与时间序列数据相关联。

注意： 如果数据集缺少完成任务所需的变量类型，则会禁用任务。例如，如果数据集不包含字符串变量，则从字符串创建日期/时间变量的任务不适用，因而被禁用。

IBM SPSS Statistics 中的日期和时间

IBM® SPSS® Statistics 中表示日期和时间的变量为数值变量，显示格式对应于特定的日期/时间格式。这些变量统称日期/时间变量。将实际表示日期的日期/时间变量与表示独立于任何日期的持续时间的变量（如 20 小时、10 分钟和 15 秒钟）区分。后者称为期间变量，前者称为日期或日期/时间变量。有关显示格式的完整列表，请参见命令语法参考的“通用”部分中的“日期和时间”。

日期和日期/时间变量。 日期变量具有表示日期的格式，如 mm/dd/yyyy。日期/时间变量具有表示日期和时间的格式，如 dd-mmm-yyyy hh:mm:ss。在内部，日期和日期/时间变量存储为从 1582 年 10 月 14 日开始的秒数。日期和日期/时间变量有时也称为日期-格式变量。

- 两位和四位数字年份规格都已经识别。缺省情况下，两位数字年份表示从当前日期的前 69 年到后 30 年这一范围。这一范围由您的选项设置确定且可被设定（从编辑菜单中，选择**选项**并单击**日期**选项卡）。
- 破折号、句号、逗号、斜线或空格都可在日-月-年格式中用作分隔符。
- 月份可以用数字、罗马数字或三个字符的缩写形式表示，也可以使用全拼的格式。三个字母的缩写和全拼的月份名称必须使用英文；使用其它语言的月份名称无法被识别。

持续时间变量。 持续时间变量具有一个表示持续时间的格式，如 hh:mm。 它们在内部存储为没有参照特定日期的秒数。

- 在时间规格中，（适用于日期/时间和持续时间变量），冒号可用作小时、分钟和秒数之间的分隔符。必须有小时和分钟，但秒有没有都可以。必须用句号分隔秒数与小数秒数。小时数可无限度大，但分钟数最大值为 59 而秒数为 59.999...。

当前日期和时间。 系统变量 $TIME 包括当前日期和时间。它表示从 1582 年 10 月 14 日到使用它的转换命令执行时的日期和时间期间的秒数。

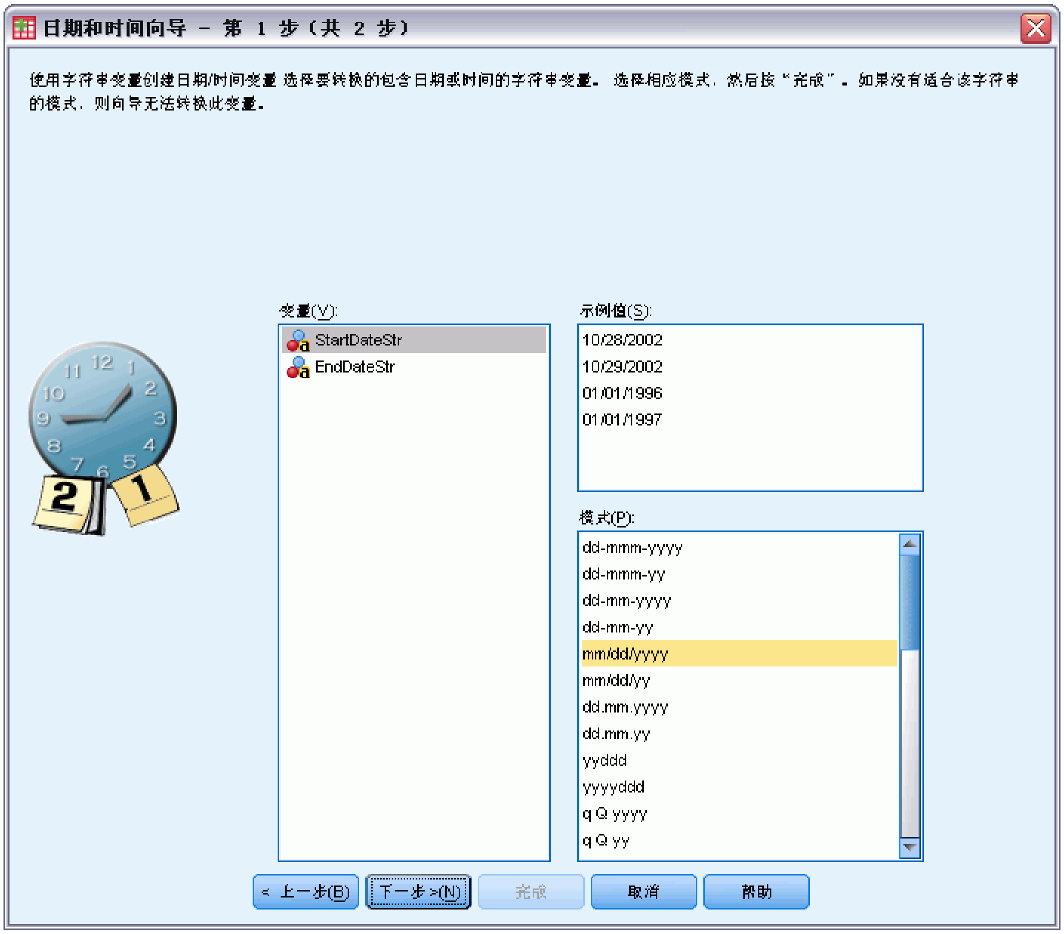
从字符串中创建一个日期/时间变量

从字符串变量中创建一个日期/时间变量：

▶ 在日期和时间向导简介屏幕上选择**从包含日期或时间的字符串中创建一个日期/时间变量**。

选择字符串变量转换至日期/时间变量

图片 8-17
从字符串变量中创建日期/时间变量，第 1 步

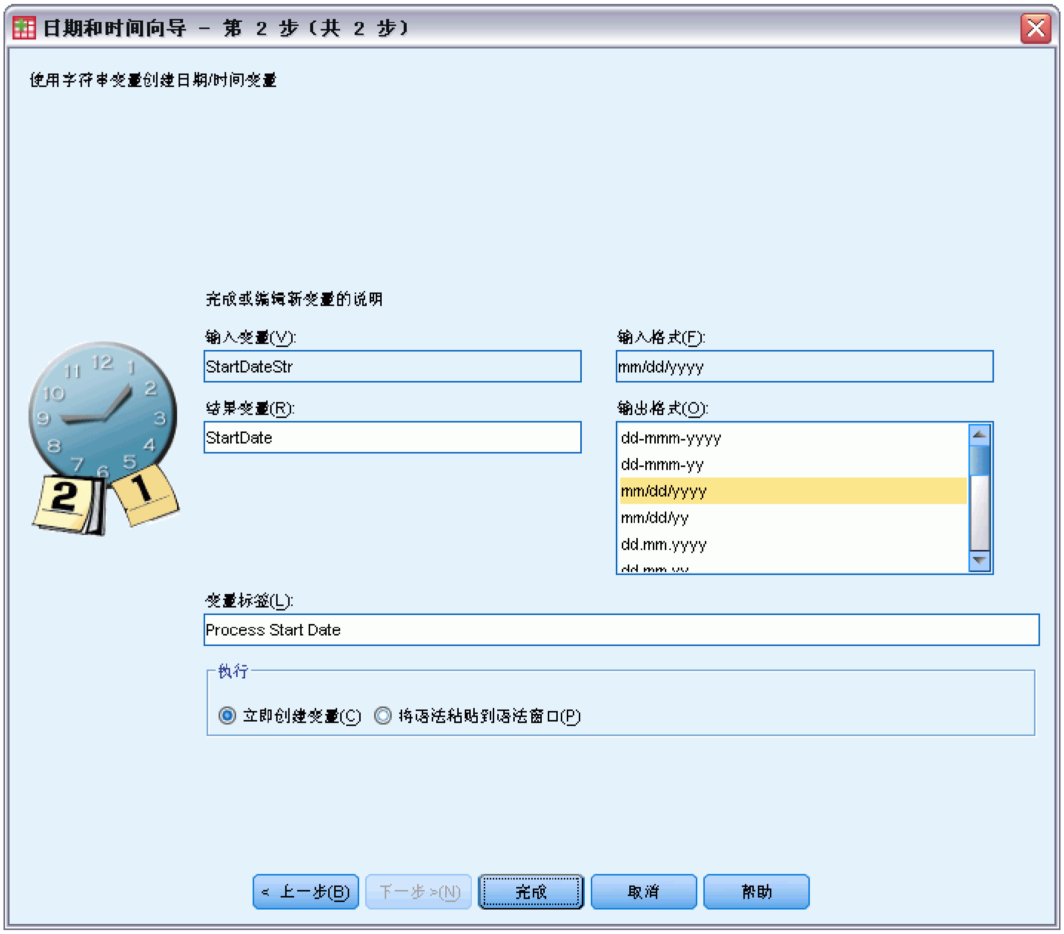


▶ 选择字符串变量在变量列表转换。注意，该列表只显示字符串变量。

▶ 从与通过字符串变量表示的日期格式匹配的模式列表中选择模式。样本值列表在数据文件夹中显示所选变量的实际值。不适用于所选模式的字符串变量值将导致新变量的系统缺失值。

指定将字符串变量转换至日期/时间变量的结果

图片 8-18
从字符串变量中创建日期/时间变量，第 2 步



▶ 为结果变量输入名称。不得为现有变量的名称。

根据需要，您可以：

■ 从输出格式列表中为新变量选择一个日期/时间格式。

■ 为新变量指定一个描述性变量标签。

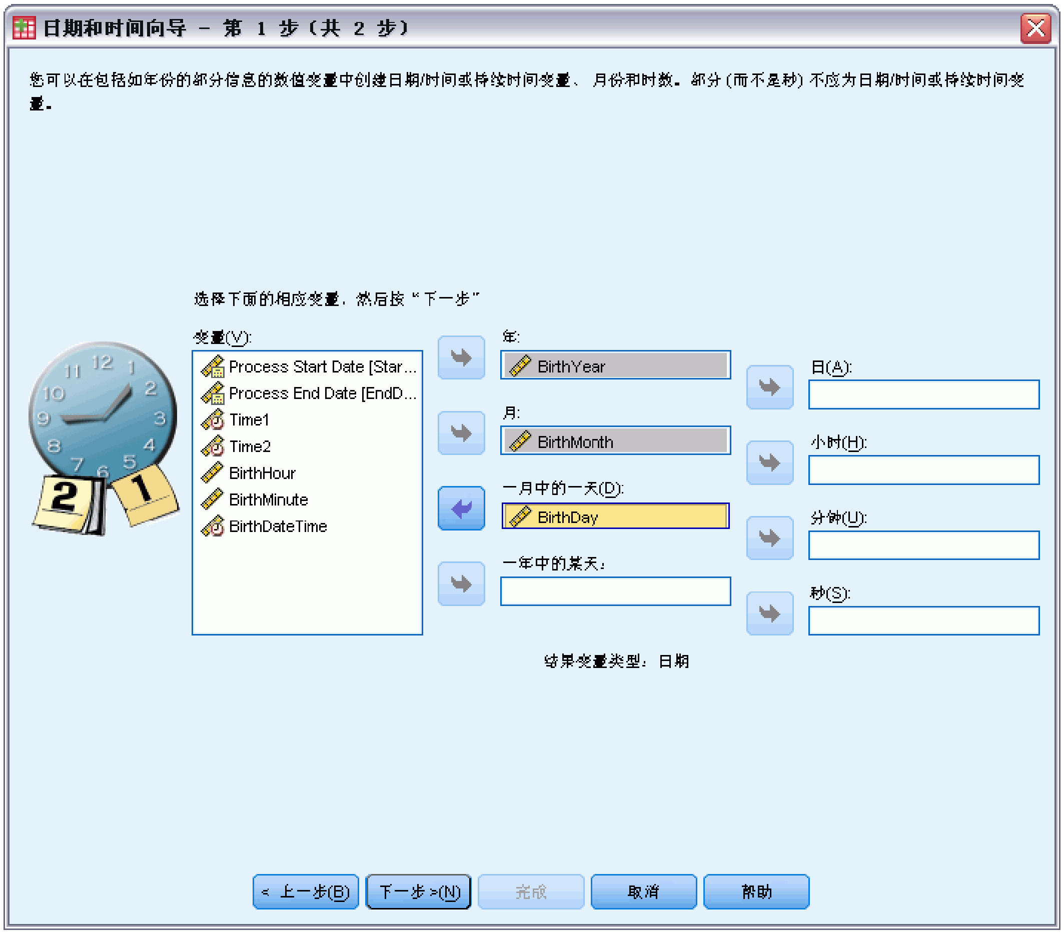
从变量组中创建一个日期/时间变量

将一组现有变量合并到单个日期/时间变量：

▶ 在日期和时间向导的介绍屏幕上选择**从包含部分日期或时间的变量中创建一个日期/时间变量**。

选择变量合并到单个日期/时间变量

图片 8-19
从变量集中创建日期/时间变量，第 1 步

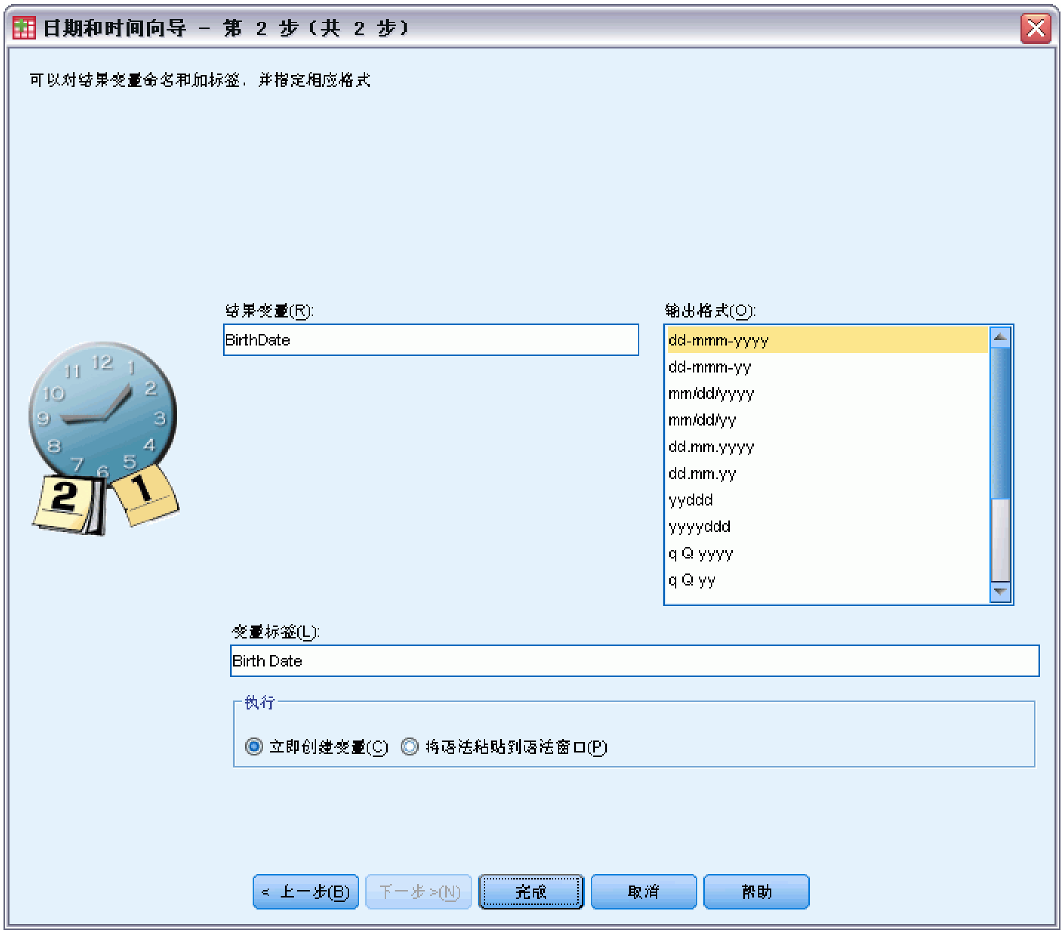


▶ 选择表示日期/时间的不同部分的变量。

- 不允许某些选择组合。例如，从年、月、日中创建一个日期/时间变量无效，因为一旦年份被选中，就需要一个完整日期。
- 您不得使用现有的日期/时间变量作为您正在创建的最终日期/时间变量的一部分。组成新日期/时间变量的变量必须为整数。允许使用现有日期/时间变量作为新变量的秒数部分是个例外。由于允许使用小数秒数，用于秒数的变量不必为整数。
- 对于新变量的任何部分，超出允许范围的值将产生新变量的系统缺失值。例如，如果您不小心使用了一个表示一月中某天的变量，由于 IBM® SPSS® Statistics 中月份有效范围是 1 – 13，将为新变量的系统缺失值指定范围 14 – 31 内任何含有一月中某天值的个案。

指定由合并变量创建的日期/时间变量

图片 8-20
从变量集中创建日期/时间变量，第 2 步



▶ 为结果变量输入名称。不得为现有变量的名称。

▶ 从输出格式列表中选择一个日期/时间格式。

根据需要，您可以：

■ 为新变量指定一个描述性变量标签。

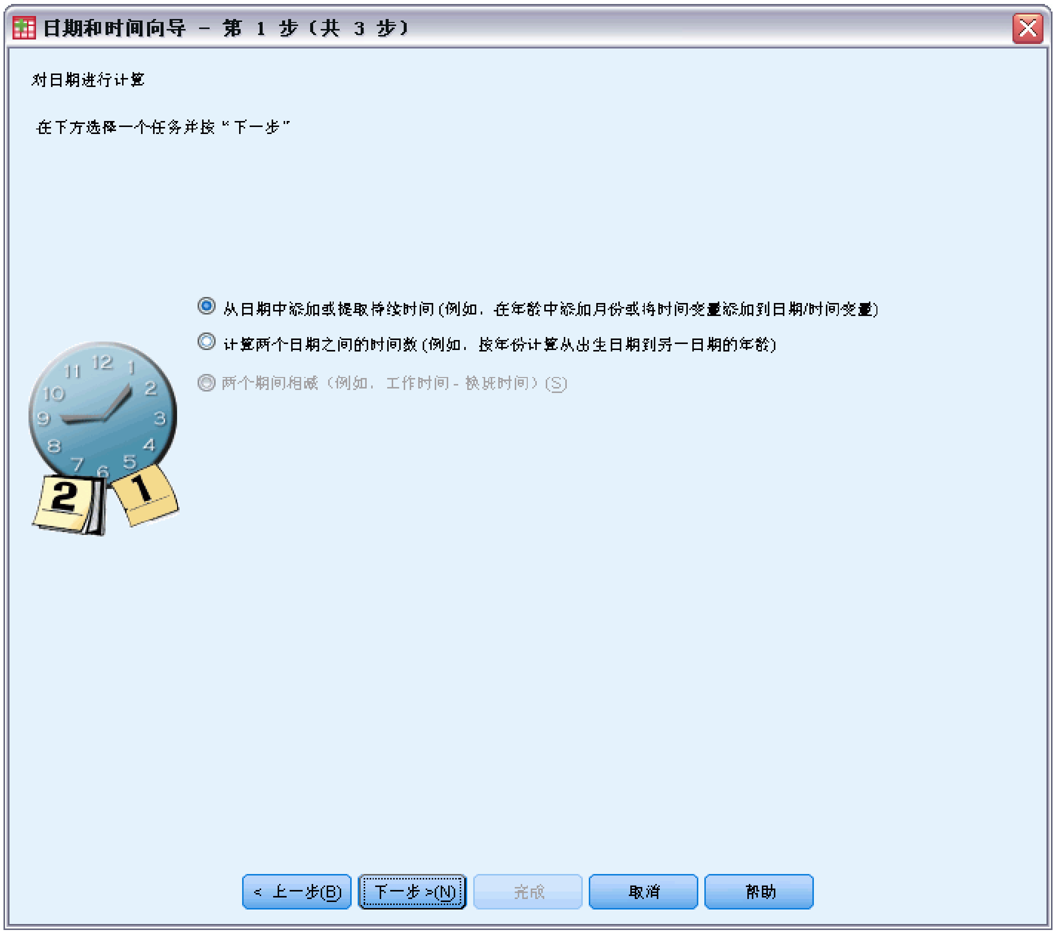
从日期/时间变量中加减值

在日期/时间变量中加减值：

▶ 在日期和时间向导的介绍屏幕上选择**使用日期和时间进行计算**。

选择计算类型以使用日期/时间变量执行

图片 8-21
从日期/时间变量中加减值，第 1 步



- **从日期中添加或提取持续时间。** 使用此选项在日期格式变量中加上或减去。您可以增加或减少为固定值的持续时间，如 10 天，或数值变量值，如表示年份的变量。
- **计算两个日期之间的时间数。** 使用此选项获取在所选单元测量的两个日期间的差别。例如，您可以获取分隔两个日期的年数或天数。
- **提取两个持续时间。** 使用此选项获取两个具有持续时间格式的变量间的差别。如 hh:mm 或 hh:mm:ss。

注意：如果数据集缺少完成任务所需的变量类型，则会禁用任务。例如，如果数据集缺少两个具有持续时间格式的变量，则提取两个持续时间的任务将不适用并被禁用。

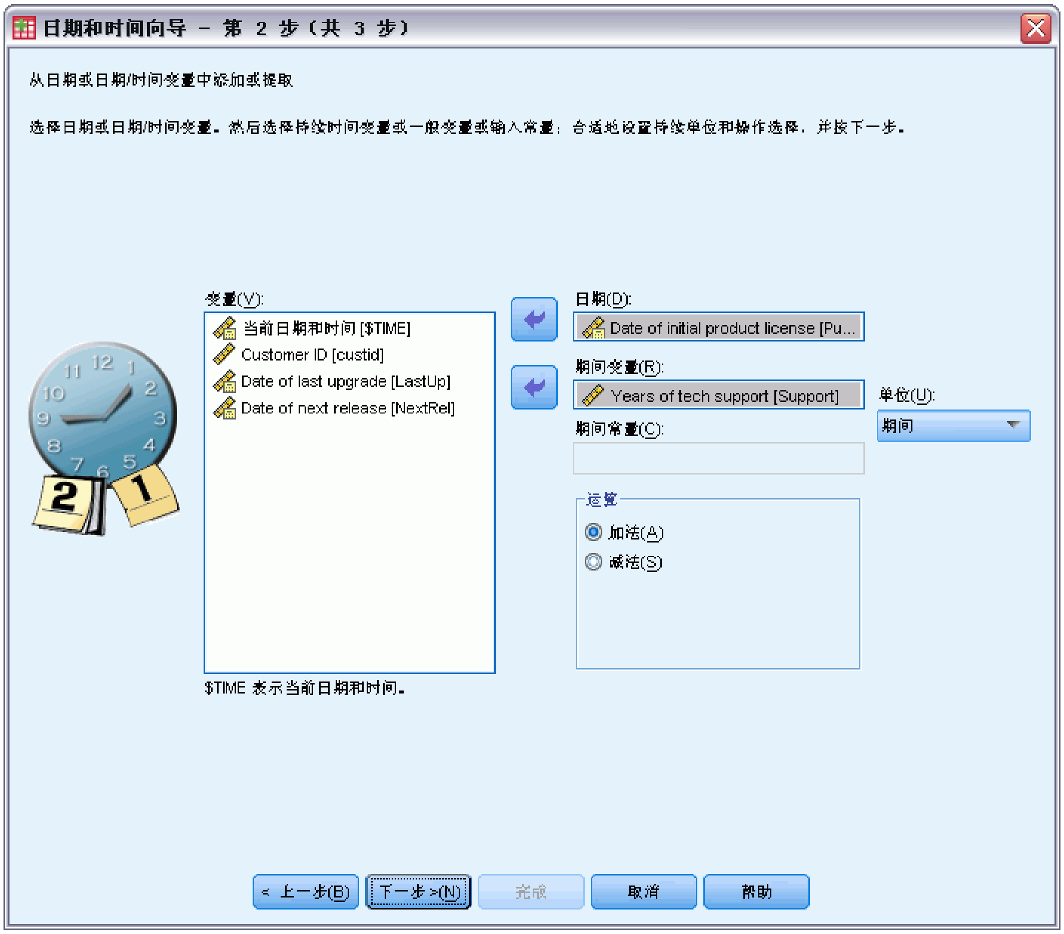
从日期中添加或提取持续时间

从日期格式变量中添加或提取持续时间：

▶ 在标有在日期上计算的日期和时间向导的介绍屏幕上选择**从日期中增加或减少持续时间**。

选择日期/时间变量和要加减的持续时间

图片 8-22
加上或减去持续时间，第 2 步



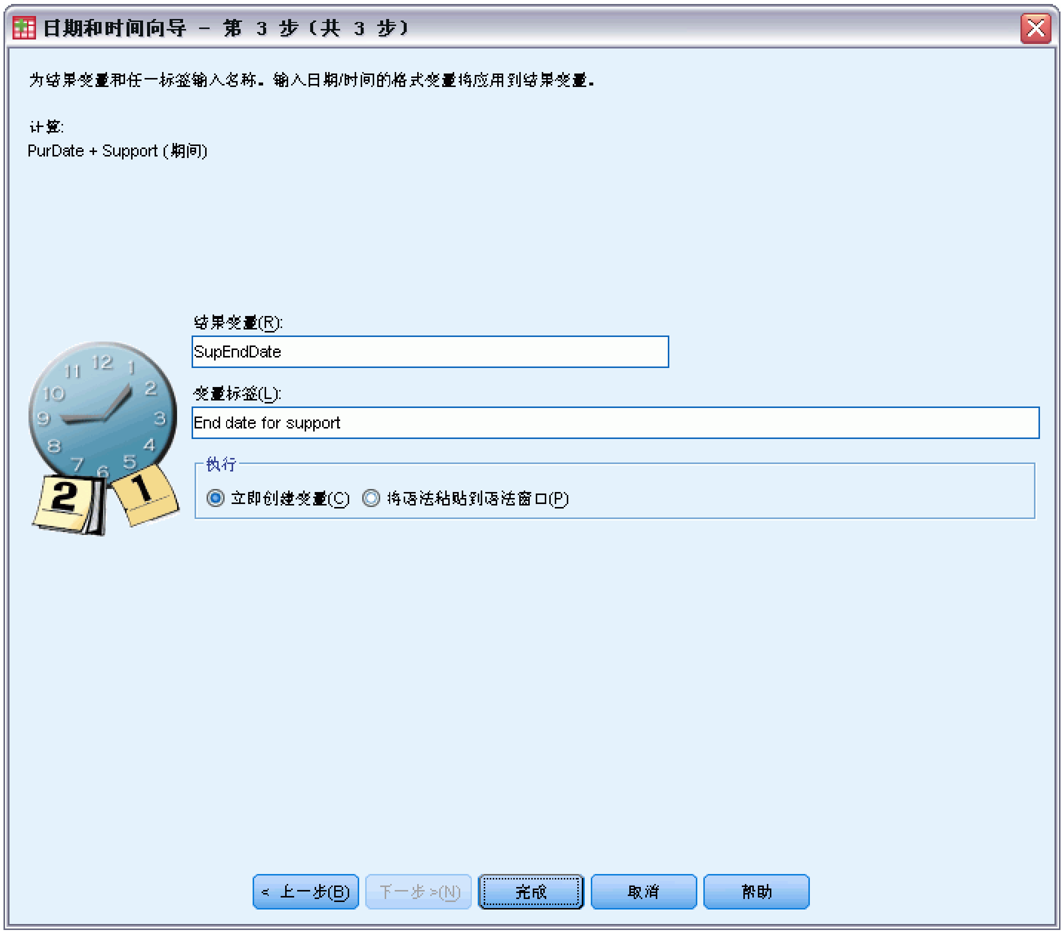
▶ 选择日期（或时间）变量。

▶ 选择一个持续时间变量或为持续时间常量输入一个值。用于持续时间的变量不得为日期或日期/时间变量。它们可以为持续时间变量或简单的数值变量。

▶ 从下拉列表中选择持续时间表示的单位。如果使用一个变量且该变量为持续时间格式，如 hh:mm 或 hh:mm:ss 时，选择**持续时间**。

从日期/时间变量中指定增加或减少持续时间的结果

图片 8-23
加上或减去持续时间，第 3 步



▶ 为结果变量输入名称。不得为现有变量的名称。

根据需要，您可以：

■ 为新变量指定一个描述性变量标签。

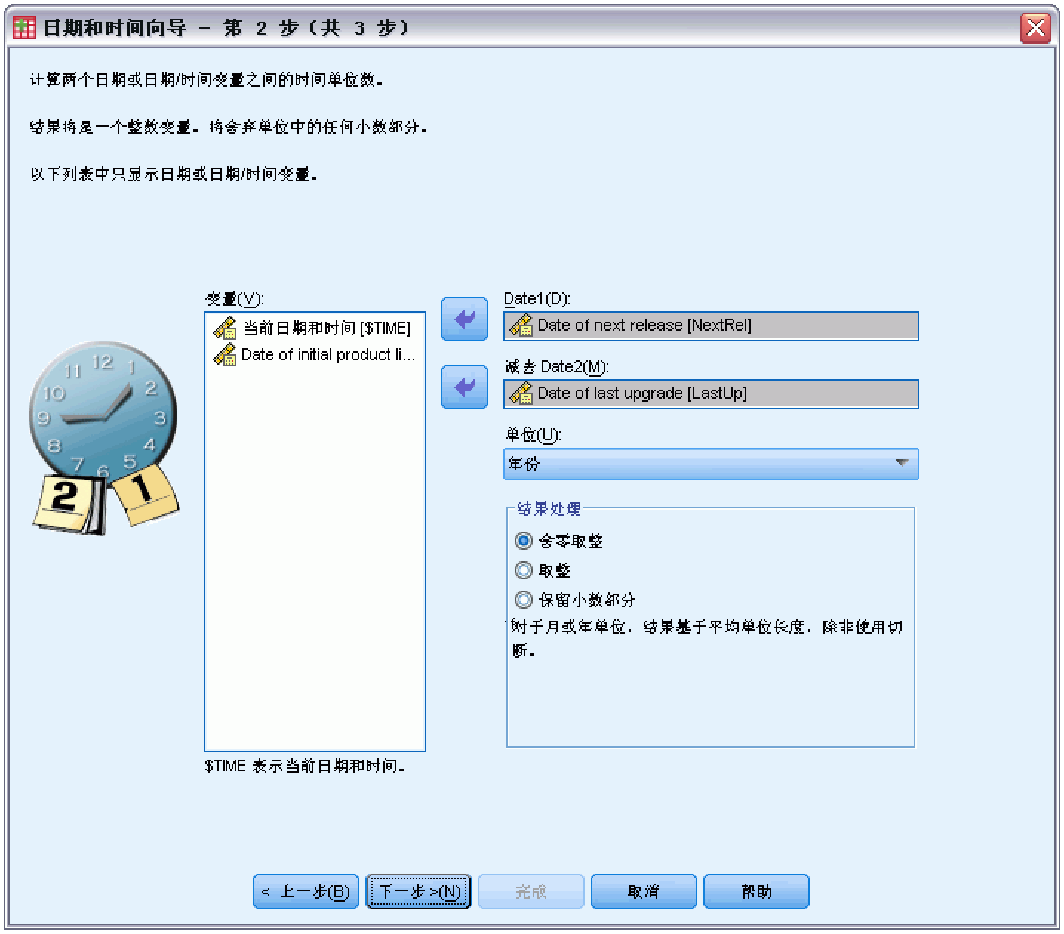
提取日期格式变量

提取两个日期格式变量：

▶ 在标有在日期上计算的日期和时间向导的介绍屏幕上选择**计算两个日期之间的时间数**。

选择要提取的日期格式变量

图片 8-24
提取日期，第 2 步



▶ 选择要提取的变量。

▶ 从下拉列表中为结果选择单位。

▶ 选择计算结果的方法（结果处理）。

结果处理

以下选项可用于计算结果的方法：

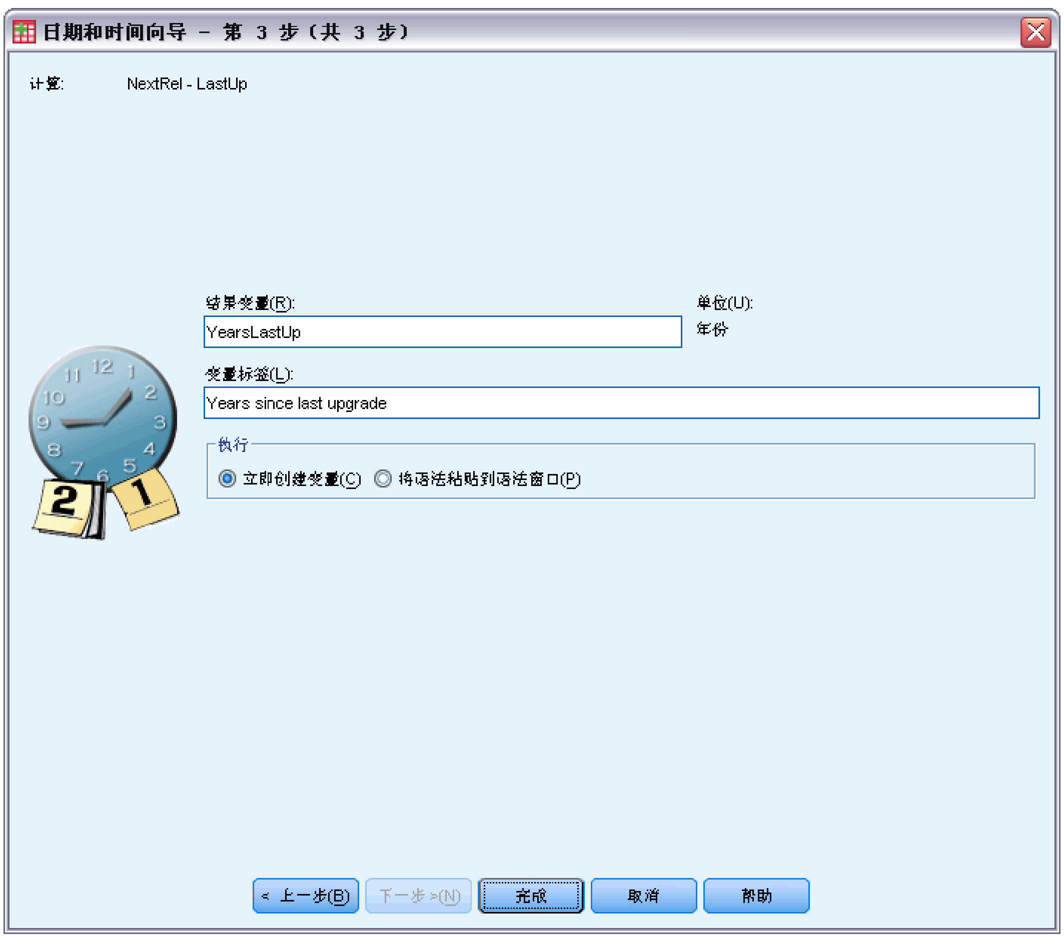
- **舍零取整。** 忽略结果的任何小数比例。例如，从 10/21/2007 中提取 10/28/2006 返回年份为 0 和月份为 11 的结果。
- **取整。** 结果四舍五入为最接近的整数。例如，从 10/21/2007 中提取 10/28/2006 返回年份为 1 和月份为 12 的结果。
- **保留小数部分。** 保留完整值；任何舍入或取整都不适用。例如，从 10/21/2007 中提取 10/28/2006 返回年份为 .98 和月份为 11.76 的结果。

为保留舍入和小数，年份结果基于一年中平均天数（365.25），而月份基于一月中平均天数（30.4375）。例如，从 3/1/2007（m/d/y 格式）中提取 2/1/2007 返回月份为 0.92 的小数结果，反之从 2/1/2007 中提取 3/1/2007 返回月份为 1.02 的小数差。这同样会影响到包括闰年在内的时间跨度的计算值。例如，与在非闰年的相同时间跨度 0.92 相比，从 3/1/2008 中提取 2/1/2008 返回一个月份为 0.95 的小数差。

日期 1	日期 2	年份			月份		
		取整	舍入	小数	取整	舍入	小数
10/21/2006	10/28/2007	1	1	1.02	12	12	12.22
10/28/2006	10/21/2007	0	1	.98	11	12	11.76
2/1/2007	3/1/2007	0	0	.08	1	1	.92
2/1/2008	3/1/2008	0	0	.08	1	1	.95
3/1/2007	4/1/2007	0	0	.08	1	1	1.02
4/1/2007	5/1/2007	0	0	.08	1	1	.99

指定提取两个日期格式变量的结果

图片 8-25
提取日期，第 3 步



▶ 为结果变量输入名称。不得为现有变量的名称。

根据需要，您可以：

■ 为新变量指定一个描述性变量标签。

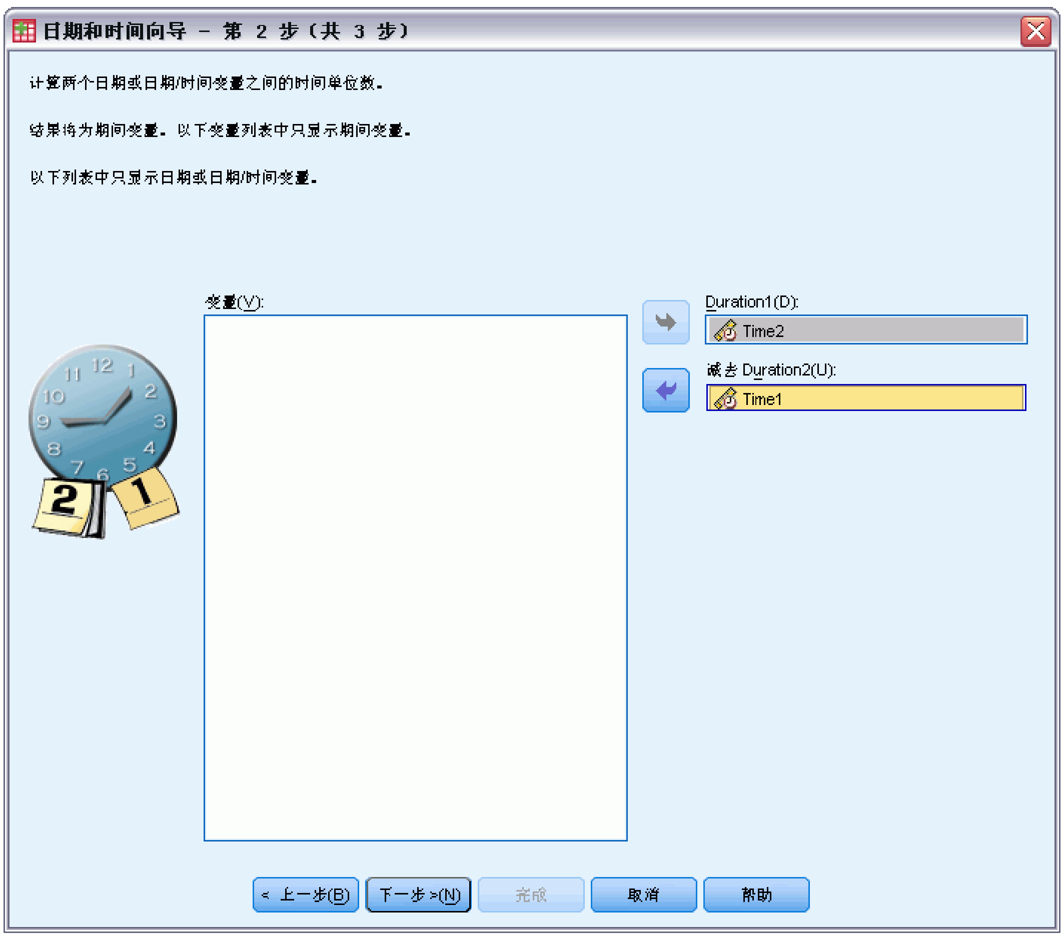
提取持续时间变量

提取两个持续时间变量：

▶ 在标有在日期上计算的日期和时间向导屏幕上选择**提取两个持续时间**。

选择要提取的持续时间变量

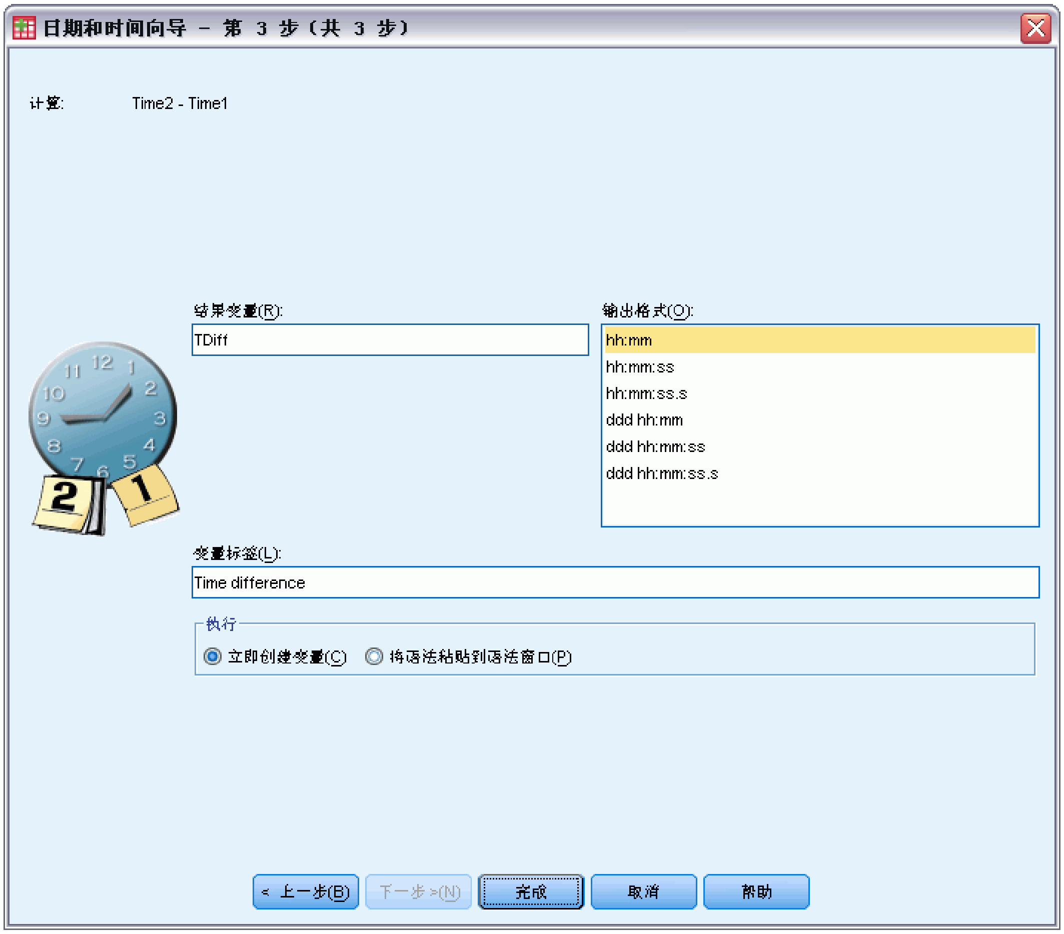
图片 8-26
提取两个持续时间，第 2 步



▶ 选择要提取的变量。

指定提取两个持续时间变量的结果

图片 8-27
提取两个持续时间，第 3 步



▶ 为结果变量输入名称。不得为现有变量的名称。

▶ 从输出格式列表中选择一个持续时间格式。

根据需要，您可以：

■ 为新变量指定一个描述性变量标签。

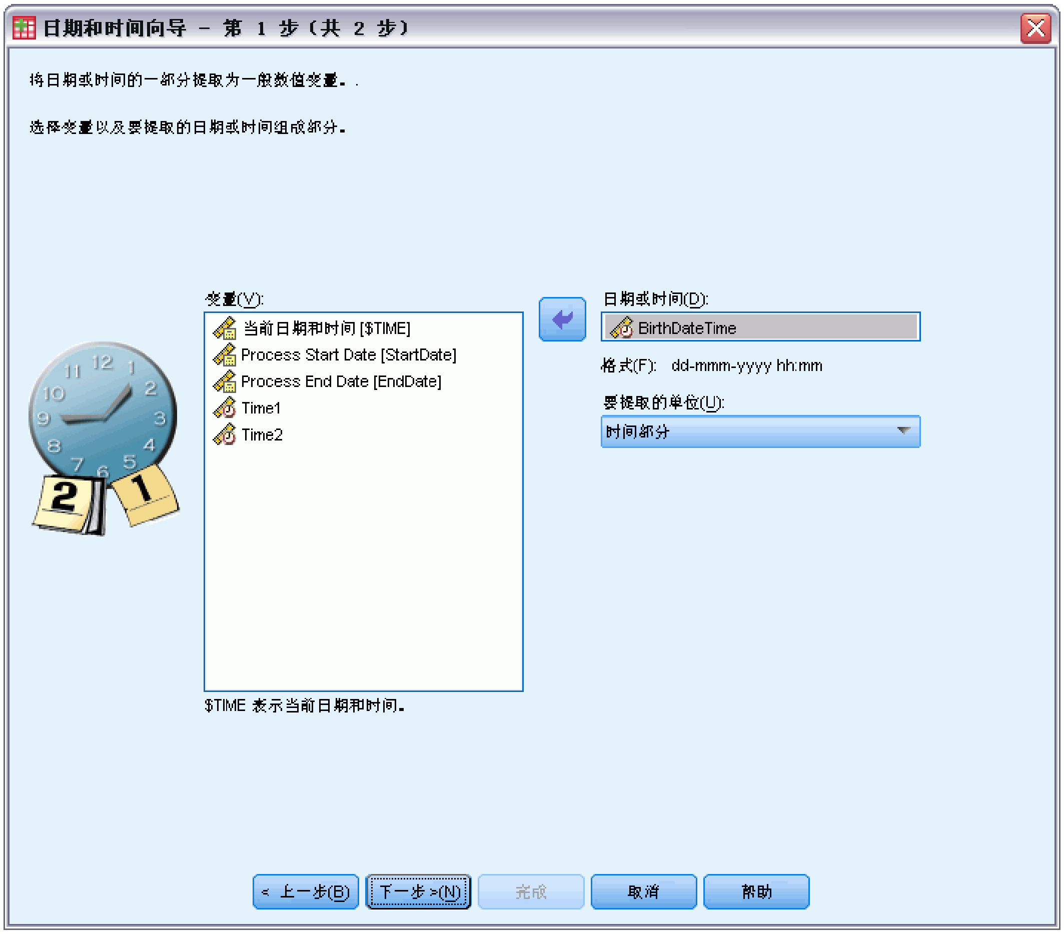
提取部分日期/时间变量

从一个日期/时间变量中提取一个成分，如年份：

▶ 在日期和时间向导的介绍屏幕上选择**提取一个日期或时间变量的一部分**。

从日期/时间变量中选择要提取的成分

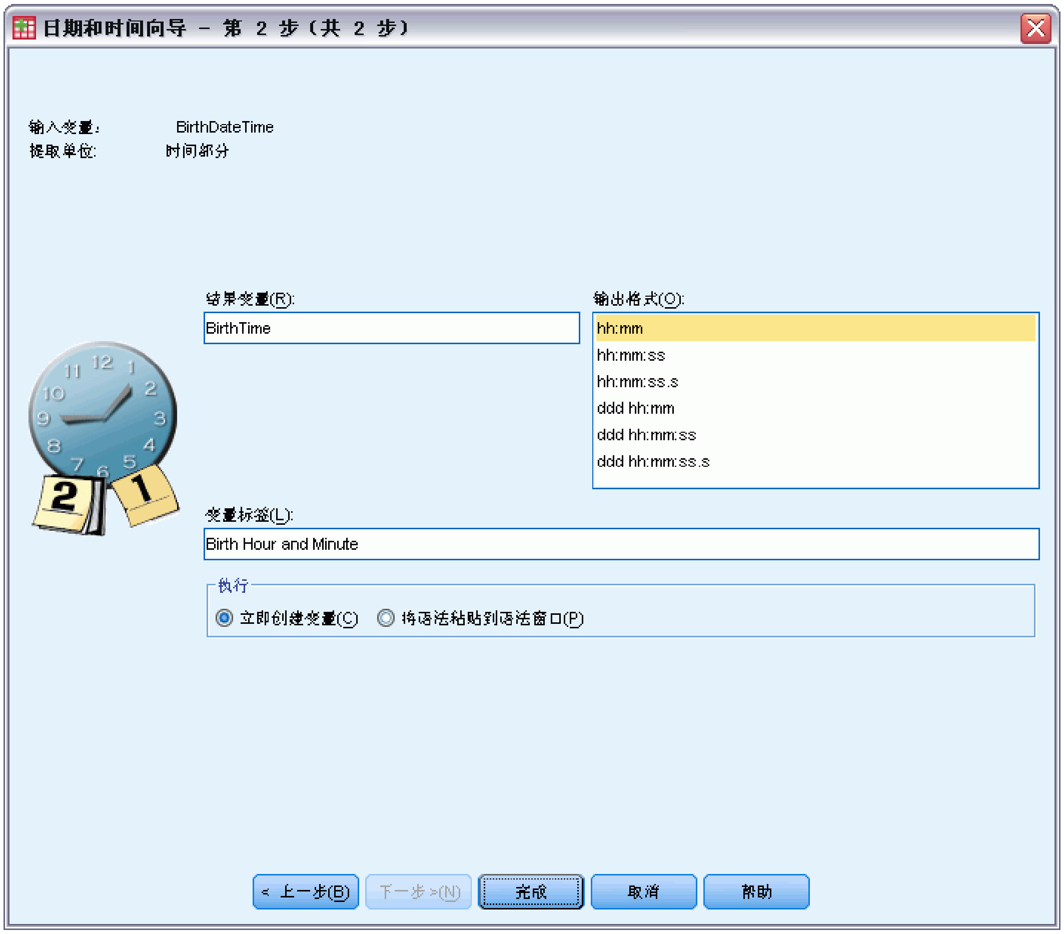
图片 8-28
获取部分日期/时间变量，第 1 步



- 选择要提取的包含日期或时间部分的变量。
- 从下拉列表中选择要提取的部分变量。您可以从明确不是显示日期部分的日期中提取信息，如一周中的一天。

指定从日期/时间变量中提取成分的结果

图片 8-29
获取部分日期/时间变量，第 2 步



▶ 为结果变量输入名称。不得为现有变量的名称。

▶ 如果您正要提取日期/时间变量的日期或时间部分，则您必须从输出格式列表中选择一种格式。一旦不需要输出格式，输出格式列表将被禁用。

根据需要，您可以：

■ 为新变量指定一个描述性变量标签。

时间序列数据转换

提供了一些在时间序列分析中有用的数据转换：

■ 生成日期变量来建立周期性并区分历史周期、验证周期和预测周期。

■ 创建新的时间序列变量作为现有时间序列变量的函数。

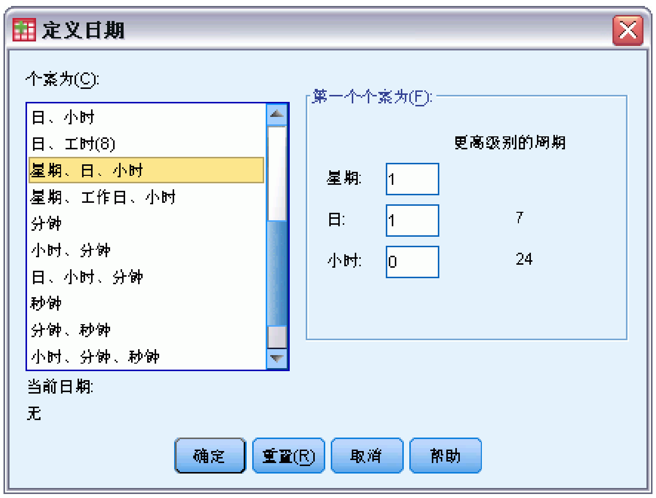
■ 使用基于一个方法的估计值替换系统缺失值和用户缺失值。

时间序列可通过定期度量某个变量（或变量集）来获得。时间序列数据转换假定这样一种数据文件结构：其中每个个案（行）代表不同时间的一组观察值，而个案之间的时间长度是均匀的。

定义日期

使用“定义日期”对话框可以生成日期变量，用于建立**时间序列**周期性和标注来自时间序列分析的输出。

图片 8-30
“定义日期”对话框



个案为。定义用于生成日期的时间区间。

- **未注日期**将移去以前定义的所有日期变量。将删除带有以下名称的任何变量：year_、quarter_、month_、week_、day_、hour_、minute_、second_ 和 date_。
- **自定义**表示存在由命令语法创建的自定义日期变量（例如四天制工作周）。此项仅反映活动数据集的当前状态。在列表中选择它不会有任何影响。

第一个个案为。定义指定给第一个个案的起始日期值。基于时间区间的序列值将指定给后面的个案。

较高级别的周期性。表示重复性循环变动，例如一年中的月数或者一周中的天数。显示的值表示您可以输入的最大值。对于小时、分钟和秒数，最大值为显示值减去 1。

将为用于定义日期的每个成分创建新的数值变量。新变量名称以下划线结尾。还将从成分中创建一个描述性字符串变量 date_。例如，如果您选择**周、天、小时**，将创建四个新变量：week_、day_、hour_ 和 date_。

如果已经定义了日期变量，当您定义的新日期变量与现有日期变量同名时，现有日期变量将被替换。

为时间序列数据定义日期

▶ 从菜单中选择：

数据 > 定义日期...

- 从“个案为”列表中选择一个时间区间。
- 输入定义“第一个个案为”的起始日期的值，该值将确定指定给第一个个案的日期。

日期变量与日期格式变量

使用“定义日期”创建的日期变量不应与数据编辑器的“变量视图”中定义的日期格式变量混淆。日期变量用于为时间序列数据创建周期性。日期格式变量代表以各种日期/时间格式显示的日期和/或时间。日期变量是简单的整数，代表距离用户指定的起始点的天数、周数、小时数等。多数日期格式变量在内部存储为自 1582 年 10 月 14 日以来的累计秒数。

创建时间序列

“创建时间序列”对话框允许您基于现有数值型**时间序列**变量的函数创建新的变量。这些转换后的值在时间序列分析中非常有用。

新变量默认名称为用来创建它的现有变量的前六个字符，然后紧跟一个下划线和序列号。例如，对于变量 price，新变量名称可以是 price_1。新变量保留来自原始变量的已定义值标签。

创建时间序列的可用函数包括差分、移动平均数、移动中位数、延迟和提前函数。

图片 8-31
“创建时间序列”对话框



要创建新的时间序列变量

- 从菜单中选择：
  转换 > 创建时间序列...

▶ 选择要用来转换原始变量的时间序列函数。

▶ 选择要从中创建新的时间序列变量的变量。仅可使用数值变量。

根据需要，您可以：

- 输入变量名以覆盖默认的新变量名。
- 为选定变量更改函数。

时间序列转换函数

差分。序列中相邻值之间的非季节性差异。阶数为用于计算差分的以前值个数。由于每阶差分丢失一个观察值，因此系统缺失值会出现在序列开头。例如，如果差分阶数为 2，则值，则前两个个案会包含新变量的系统缺失值。

季节性差分。相隔恒定距离的序列值之间的差分。该跨度基于当前定义的周期。要计算季节性差分，您必须定义日期变量（“日期”菜单，“定义日期”），其中包括周期性成分（例如一年中的月份）。阶数为用于计算差分的季节性周期个数。在序列开头，带有系统缺失值的个案个数，等于阶数乘以周期。例如，如果当前周期为 12，且阶数为 2，则前 24 个个案会包含新变量的系统缺失值。

中心移动平均数。当前序列值与其周围某个跨度内序列值的平均值。跨度为用于计算平均值的序列值个数。如果跨度为偶数，则移动平均数通过对每组非中心平均值求平均值而得出。在跨度为 n 的序列开头和末尾，带有系统缺失值的个案个数，等于 n/2（偶数跨度值）和 (n－1)/2（奇数跨度值）。例如，如果跨度为 5，则在序列开头和末尾带有系统缺失值的个案个数为 2。

前移动平均数。当前序列值之前的序列值的平均值。跨度为用于计算平均值的前面序列值个数。在序列开头，带有系统缺失值的个案个数，等于跨度值。

移动中位数。当前序列值与其周围某个跨度内序列值的中位数。跨度为用于计算中位数的序列值个数。如果跨度为偶数，则中位数通过对每组非中心中位数求平均值而得出。在跨度为 n 的序列开头和末尾，带有系统缺失值的个案个数，等于 n/2（偶数跨度值）和 (n－1)/2（奇数跨度值）。例如，如果跨度为 5，则在序列开头和末尾带有系统缺失值的个案个数为 2。

累积和。当前序列值与其周围序列值的累积和。

延迟。根据指定的延迟阶数，上一个个案的值。阶数为从中获取值的当前个案之前的个案个数。在序列开头，带有系统缺失值的个案个数，等于阶数值。

提前。根据指定的提前阶数，后一个个案的值。阶数为从中获取值的当前个案之后的个案个数。在序列末尾，带有系统缺失值的个案个数，等于阶数值。

平滑。基于复合数据平滑器的新序列值。平滑器从移动中位数 4 开始，由移动中位数 2 居中。然后，它再通过移动中位数 5、移动中位数 3 和 Hanning 加权平均，重新对这些值进行平滑。从原始序列中减去平滑后的序列，计算得出残差。然后对计算得出的残差重复这整个过程。最后，减去该过程首次获得的平滑值，得到平滑残差。这有时也称为 **T4253H 平滑**。

替换缺失值

缺失观察值可能会在分析中导致问题，如果序列中缺少值，会无法计算某些时间序列度量。有时，仅仅是特定的观察值未知。此外，数据缺失的原因可能包括：

- 每个差分度会使序列长度减 1。
- 每个季节性差分度会使序列长度减少 1 个季节。
- 如果您创建的新序列包含在现有序列末尾之外的预测（单击**保存**按钮，选择合适选项），原始序列和产生的残差序列会缺少新观察值的数据。
- 某些转换（例如，对数转换）会导致缺失原始序列的特定值数据。

序列开头和末尾的缺失数据不会引发特殊的问题，只会缩短序列的有效长度。序列中间的缺口（内嵌缺失数据）是更为严重的问题。根据使用的分析过程，问题程度也有所不同。

“替换缺失值”对话框允许您从现有变量创建新的**时间序列**变量，将缺失值替换为根多种方法之一估计的值。新变量默认名称为用来创建它的现有变量的前六个字符，然后紧跟一个下划线和序列号。例如，对于变量 price，新变量名称可以是 price_1。新变量保留来自原始变量的已定义值标签。

图片 8-32
“替换缺失值”对话框



要替换时间序列变量的缺失值

▶ 从菜单中选择：
转换 > 替换缺失值...

▶ 选择要用于替换缺失值的估计方法。

▶ 选择要替换其缺失值的变量。

根据需要，您可以：

- 输入变量名以覆盖默认的新变量名。
- 为选定变量更改估计方法。

替换缺失值的估计方法

序列平均值。使用整个序列的平均值替换缺失值。

邻近点的平均值。使用有效周围值的平均值替换缺失值。邻近点的跨度为缺失值上下用于计算平均值的有效值个数。

邻近点的中位值。使用有效周围值的中位值替换缺失值。邻近点的跨度为缺失值上下用于计算中位值的有效值个数。

线性插值。使用线性插值替换缺失值。缺失值之前的最后一个有效值和之后的第一个有效值用来作为插值。如果序列中的第一个或最后一个个案具有缺失值，则不必替换。

该点的线性趋势。使用该点的线性趋势替换缺失值。现有序列在标度为从 1 到 n 的索引变量上回归。采用其预测值替换缺失值。

章

9

文件处理和文件转换

数据文件的组织形式对于特定需要而言并非总是理想的。您可能想要合并数据文件、以不同的顺序对数据排序、选择个案的子集、或通过对个案进行分组来更改分析的单位。有大量可用的文件转换功能，包括：

对数据排序。您可根据一个或多个变量的值对个案排序。

转置个案和变量。IBM® SPSS® Statistics 数据文件格式将行作为个案读取，将列作为变量读取。对于顺序相反的数据文件，您可交换行和列并以正确的格式读取数据。

合并文件。您可合并两个或更多个数据文件。您可将变量相同但个案不同或个案相同但变量不同的文件进行合并。

选择个案的子集。您可以将分析限制为个案的子集或对不同的子集执行同步分析。

对数据分类汇总。您可通过基于一个或多个分组变量的值对个案分类汇总而更改分析的单位。

对数据加权。您可以基于权重变量的值对个案加权以进行分析。

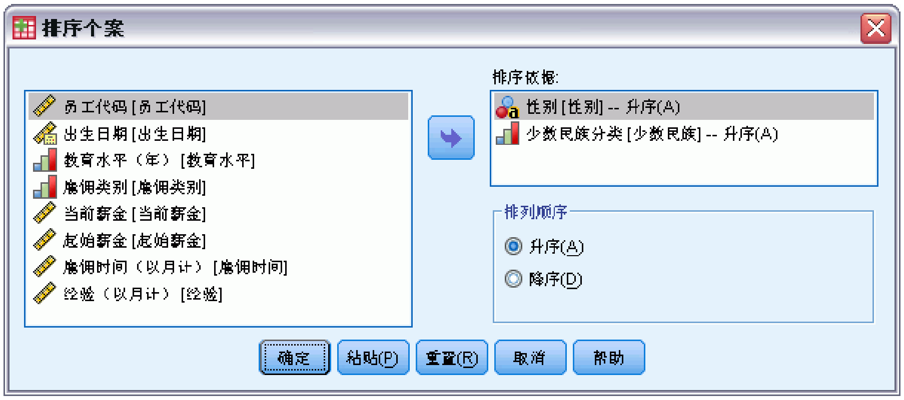
重组数据。您可重组数据以从多个个案中创建单个个案（记录），或从单个个案中创建多个个案。

排序个案

此对话框基于一个或多个排序变量的值对数据文件的个案（行）排序。您可以按升序或降序对个案排序。

- 如果您选择了多个排序变量，个案按“排序”列表的前一个变量的类别中的每个变量排序。例如：如果您将选择性别作为第一个排序变量，选择少数民族作为第二个排序变量，则将按每个性别类别中的少数民族分类对个案排序。
- 排序顺序基于语言环境定义的顺序（不一定与字符代码的数值顺序相同）。缺省的语言环境是操作系统的语言环境。可以控制语言环境，方法是使用“选项”对话框的“常规”选项卡上的“语言”设置（“编辑”菜单）。

图片 9-1
“个案排序”对话框



要对个案排序，请执行下列操作

▶ 从菜单中选择：
数据 > 排序个案...

▶ 选择一个或多个排序变量。

排列变量

您可以根据任何变量属性值（例如，变量名称、数据类型、测量级别），包括定制变量属性，对活动数据集的变量进行排序。

- 可以按升序或降序对值排序。
- 您可以在定制变量属性中保存原始（预先分类）变量顺序。
- 根据定制变量属性的值进行排序，只限于当前显示在变量视图中的定制变量属性。

有关定制变量属性的更多信息，请参见定制变量属性。

要排列变量

在数据编辑器的变量视图中：

▶ 右键单击属性列标题，从上下文菜单中选择**按升序排列**或**按降序排列**。

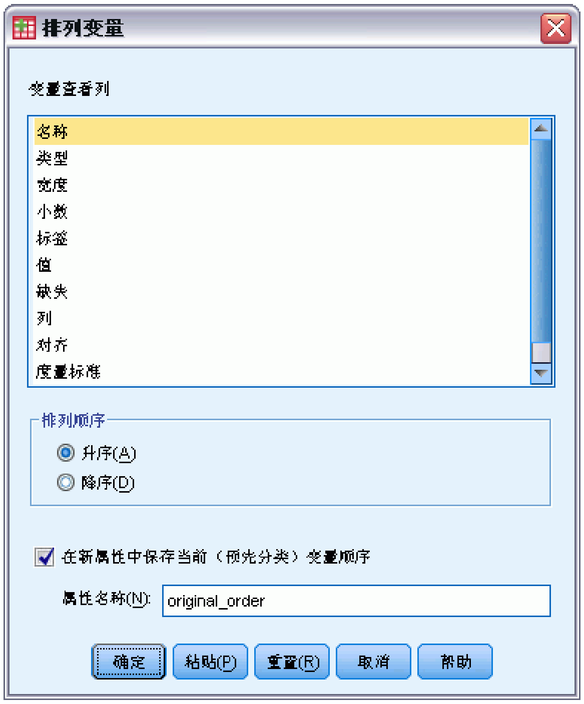
或

▶ 在“变量视图”或“数据视图”中，从菜单中选择：
数据 > 排列变量

▶ 选择用于排序变量的属性。

▶ 选择排列顺序（升序或降序）。

图片 9-2
排列变量对话框



- 变量属性列表与数据编辑器的“变量视图”中显示的属性列名称相匹配。
- 您可以在定制变量属性中保存原始（预先分类）变量顺序。对于每个变量，该属性值是一个整数值，表示它在排序前的位置；因此，在根据该定制属性值对变量进行排序时，您可以恢复原来的变量顺序。

转置

“转置”创建一个新的数据文件，初始数据文件中的行列被转置，使个案（行）成为变量同时变量（列）成为个案。“转置”自动创建新的变量名称，并显示新的变量名称列表。

- 自动创建包含初始变量名称 case_lbl 的新字符串变量。
- 如果活动数据集包含具有唯一值的 ID 或名称变量，则可以使用其作为名称变量，且其值在转置后的数据文件中将用作变量名称。如果这是数值变量，变量名称则以字母 V 开头，后跟该数值。
- 在转置后的数据文件中，用户缺失值会转换为系统缺失值。要保留所有这些值，请在“数据编辑器”的“变量视图”中更改缺失值的定义。

要转置变量和个案，请执行下列操作

▶ 从菜单中选择：

数据 > 转置...

▶ 选择一个或多个要转置为个案的变量。

合并数据文件

可以使用两种不同的方式合并两个文件中的数据。您可以：

- 合并活动数据集和另一个打开的数据集，或包含相同变量但不同个案的 IBM® SPSS® Statistics 数据文件。
- 合并活动数据集和另一个打开的数据集，或包含相同个案但不同变量的 SPSS Statistics 数据文件。

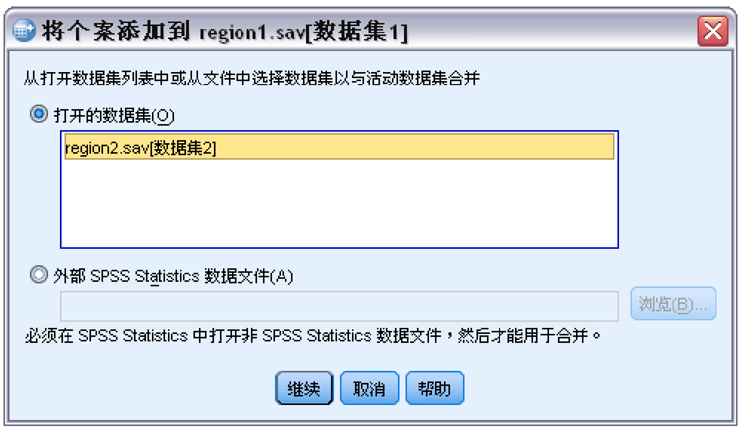
合并文件

▶ 从菜单中选择：

数据 > 合并文件

▶ 选择**添加个案**或**添加变量**。

图片 9-3
选择要合并的文件



添加个案

“添加个案”将活动数据集与另外一个数据集或包含相同变量（列）但不同个案（行）的外部 IBM® SPSS® Statistics 数据文件合并在一起。例如，您可能在两个不同的销售区域中记录客户的相同信息，并以单独文件的形式维护每个区域的数据。另外的数据集可以是外部 SPSS Statistics 数据文件，也可以是当前会话中可用的数据集。

图片 9-4
“添加个案”对话框



非成对变量。要从新的合并的数据文件中排除的变量。活动数据集中的变量使用星号 (*) 标识。外部数据文件中的变量使用加号 (+) 标识。缺省情况下，此列表包含：

- 两个数据文件中的变量名互不匹配的变量。可以在不成对的变量中创建对，并将其包含在新的合并文件中。
- 在一个文件定义为数值数据而在另一文件中被定义为字符串数据的变量。数值变量不能与字符串变量合并。
- 宽度不相等的字符串变量。两个数据文件中的字符串变量的定义宽度必须相同。

新的活动数据集中的变量。要包含在新的合并的数据文件中的变量。缺省情况下，在列表上包含所有名称和数据类型（数字或字符串）都匹配的变量。

- 如果不想将某些变量包含在合并文件中，可从列表中删除这些变量。
- 包含在合并文件中的任何不成对的变量，都包含一些个案的缺失数据，而这些个案是来自于不包含该变量的文件。

将个案源表示为变量. 指示每个个案的源数据文件。对于来自活动数据集的个案，此变量具有 0 值；对于来自外部数据文件的个案，此变量具有 1 值。

合并具有相同变量和不同个案的数据文件

▶ 打开至少一个要合并的数据文件。如果打开了多个数据集，则使其中一个要合并的数据集成为活动数据集。此文件中的个案将首先出现在新的合并数据文件中。

▶ 从菜单中选择：

数据 > 合并文件 > 添加个案...

▶ 选择要与活动数据集合并的数据集或外部 SPSS Statistics 数据文件。

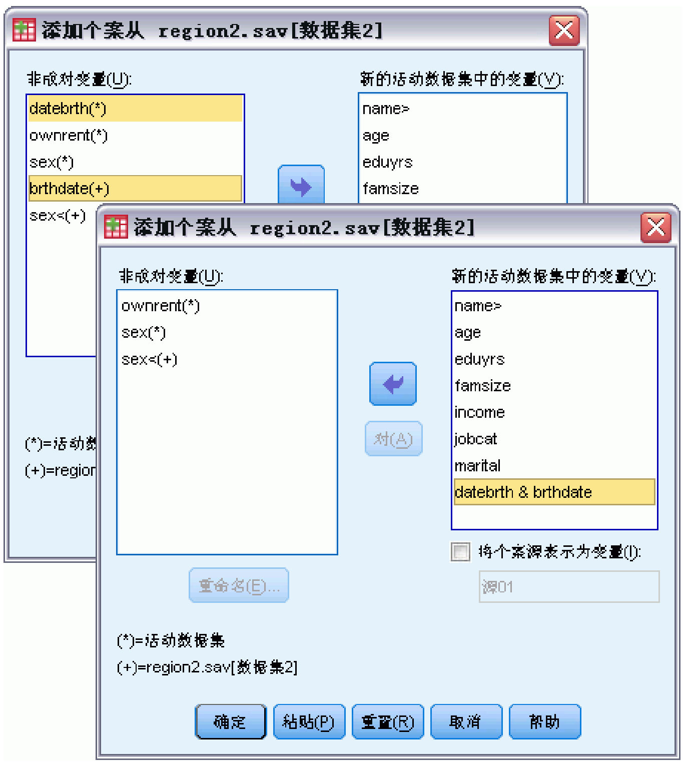
▶ 在“新活动数据集”列表的“变量”中删除不想要的任何变量。

- ▶ 任何从“不成对的变量”列表中添加的变量对都表示相同信息，这些信息以不同变量名在两个文件中记录。例如，出生日期可能在某一文件中的变量名是 brthdate，而在另一个文件中的变量名是 datebrth。

选择不成对的变量对

- ▶ 在“不成对的变量”列表中单击其中一个变量。
- ▶ 按住 Ctrl 并单击列表上的其他变量。（按 Ctrl 键并同时单击鼠标左键。）
- ▶ 单击对将变量对移到“新活动数据集”列表中的“变量”中。（活动数据集中的变量名用作合并文件中的变量名。）

图片 9-5
按住 Ctrl 键并单击来选择变量对



添加个案：重命名

可以重命名活动数据集或其他数据集中的变量，然后将这些变量从不成对列表移动到要包含在合并数据文件的变量列表中。重命名变量使您可以：

- 对于变量对，使用外部文件的变量名，而不是活动数据集中的变量名。
- 包含两个名称相同但类型不匹配或字符串宽度不同的变量。例如，要同时包含活动数据集中的数值变量性别和外部文件中的字符串变量性别，则首先必须重命名其中一个变量。

添加个案：字典信息

活动数据集中的所有现有字典信息（变量和值标签、用户缺失值、显示格式）会应用到合并数据文件中。

- 如果在活动数据集中没有定义变量的任何字典信息，则会使用另一个数据集中的字典信息。
- 如果活动数据集包含一个变量的任何已定义值标签或用户缺失值，则会忽略其他数据集中该变量的所有其他值标签或用户缺失值。

合并两个以上的数据源

使用命令语法最多可以合并 50 个数据集和/或数据文件。有关更多信息，请参见命令语法参考（可从“帮助”菜单中获得）中的 ADD FILES 命令。

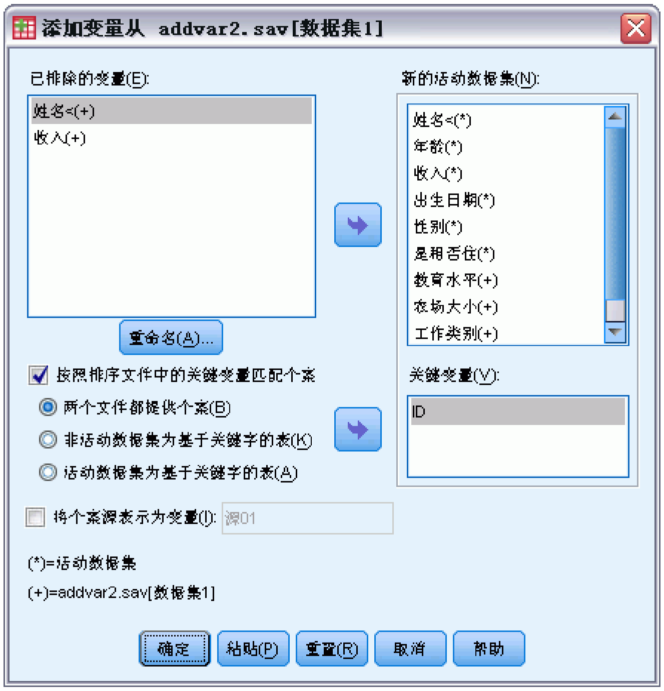
添加变量

“添加变量”将活动数据集与另一个打开的数据集或包含相同个案（行）但不同变量（列）的 IBM® SPSS® Statistics 数据文件合并在一起。例如，可能要将一个包含检验前结果的数据文件与包含检验后结果的数据文件合并。

- 在两个数据集中个案必须以相同的顺序排序。
- 如果一个或多个关键字变量用于匹配个案，就必须以关键字变量的升序顺序对这两个数据集进行排序。
- 缺省情况下，将排除第二个数据文件中与活动数据集有重复变量名的变量名，因为“添加变量”假定这些变量包含重复信息。

将个案源表示为变量. 指示每个个案的源数据文件。对于来自活动数据集的个案，此变量具有 0 值；对于来自外部数据文件的个案，此变量具有 1 值。

图片 9-6
“添加变量”对话框



已排除的变量。要从新的合并的数据文件中排除的变量。缺省情况下，此列表包含在其他数据集中与活动数据集有重复变量名的所有变量名。活动数据集中的变量使用星号 (*) 标识。外部数据文件中的变量使用加号 (+) 标识。如果您想在合并文件中包含具有重复名称的已排除变量，您可将其重命名并添加到要包含的变量列表中。

新的活动数据集。要包含在新的合并的数据集中的变量。缺省情况下，列表上包含这两个数据集中的所有唯一变量名。

关键变量。如果一个数据集中的某些个案在另一个数据集中没有匹配的个案（即：一个数据集中的某些个案缺失），请使用关键变量在两个数据集中标识并正确匹配个案。您还可将关键变量用于表查询文件。

- 关键变量在两个数据集中的名称必须相同。
- 必须以关键变量的升序对两个数据集进行排序，“关键变量”列表上的变量顺序必须与其排序顺序相同。
- 关键变量上不匹配的个案会包含在合并文件中，但不与其他文件中的个案相合并。不匹配的个案仅包含其所处原文件中的变量的值；来自其他文件中的变量包含系统缺失值。

非活动数据集或活动数据集是基于关键字的表。基于关键字的表或**表查询文件**是一种文件，其中每个“个案”的数据可以应用到其他数据文件的多个个案中。例如，如果一个文件包含每个家庭成员（如性别、年龄、教育）的信息，并且其他文件包含家庭总体信息（如总收入、家庭大小、地点），则您可将家庭数据的文件用作表查询文件，并可将公共家庭数据应用到合并数据文件中的每个个人家庭成员。

合并带有相同个案但不同变量的文件

▶ 打开至少一个要合并的数据文件。如果打开了多个数据集，则使其中一个要合并的数据集成为活动数据集。

▶ 从菜单中选择：
数据 > 合并文件 > 添加变量...

▶ 选择要与活动数据集合并的数据集或外部 SPSS Statistics 数据文件。

选择关键变量

▶ 在“排除变量”列表的外部文件变量 (+) 中选择变量。

▶ 选择**按照排序文件中的关键变量匹配个案**。

▶ 将这些变量添加到“关键变量”列表中。

关键变量必须同时存在于活动数据集和其他数据集中。必须以关键变量的升序对两个数据集进行排序，“关键变量”列表上的变量顺序必须与其排序顺序相同。

添加变量：重命名

可以重命名活动数据集或其他数据集中的变量，然后将这些变量移动到要包含在合并数据文件中的变量列表。这主要对要包含两个文件中名称相同但信息不同的两个变量很有用。

合并两个以上的数据源

使用命令语法最多可以合并 50 个数据集和/或数据文件。有关更多信息，请参见命令语法参考（可从“帮助”菜单中获得）中的 MATCH FILES命令。

分类汇总数据

“分类汇总数据”将活动数据集中的个案组汇总为单个个案并创建新的汇总文件，或在活动数据集中创建包含分类汇总数据的新变量。基于零个或多个中断（分组）变量的值汇总个案。如果未指定中断变量，则整个数据集将成为单个中断组。

- 如果创建新的分类汇总数据文件，则新数据文件对由中断变量定义的每个组都包含一个个案。例如，如果中断变量有两个值，则新的数据文件将仅包含两个个案。如果未指定中断变量，则新数据文件将包含一个个案。
- 如果将分类汇总变量添加到活动数据集，则不分类汇总数据文件本身。分隔变量值相同的每个个案对新聚合变量都得到相同的值。例如，如果性别是唯一的中断变量，则所有男性对于表示平均年龄的新分类汇总变量将得到相同的值。如果未指定中断变量，则对于代表平均年龄的新分类汇总变量，所有个案将收到相同值。

图片 9-7
“分类汇总数据”对话框



分组变量。根据中断变量的值，对个案进行分组。中断变量的每个唯一组合定义一个组。创建新的分类汇总数据文件时，所有中断变量保存在新文件中，同时保留其现有名称和字典信息。中断变量（如果已指定）可以是数值型或字符串型。

汇总变量。源变量用于分类汇总函数，以创建新的分类汇总变量。分类汇总变量名称紧跟一个可选的变量标签、分类汇总函数的名称和源变量名称（用括号括起来）。

您可使用新的变量名称覆盖缺省分类汇总变量名称，提供描述性变量标签，以及更改用于计算分类汇总数据值的函数。您还可以创建包含每个中断组中的个案数的变量。

分类汇总数据文件

- 从菜单中选择：

  数据 > 分类汇总...

- 根据需要，选择中断变量，以定义如何对个案分组以创建分类汇总数据。如果未指定中断变量，则整个数据集将成为单个中断组。

- 选择一个或多个分类汇总变量。

▶ 为每个分类汇总变量选择一个分类汇总函数。

保存分类汇总结果

可以将分类汇总变量添加到活动数据集，或创建新的分类汇总数据文件。

- **将汇总变量添加到活动数据集**. 基于分类汇总函数的新变量添加到活动数据集中。数据文件本身并不聚合。分隔变量值相同的每个个案对新聚合变量都得到相同的值。
- **创建只包含汇总变量的新数据集**. 在当前会话中将汇总数据保存到新的数据集。该文件包含定义分类汇总个案的中断变量以及由分类汇总函数定义的所有汇总变量。活动数据集不受影响。
- **写入只包含汇总变量的新数据文件**. 将汇总数据保存到外部数据文件。该文件包含定义分类汇总个案的中断变量以及由分类汇总函数定义的所有汇总变量。活动数据集不受影响。

用于大数据文件的排序选项

对于非常大的数据文件，分类汇总已预先排序的数据可能更有效。

文件已经按分组变量排序. 如果已按分隔变量的值对数据进行了排序，则该选项可以使过程更快速地运行，并且占用更少的内存。应谨慎使用该选项。

- 数据必须按照中断变量的值进行排序，且中断变量的顺序必须与为“分类汇总数据”过程指定的中断变量的顺序相同。
- 如果要将变量添加到活动数据集，则只有在按照中断变量的升序值对数据进行了排序的情况下才选择此选项。

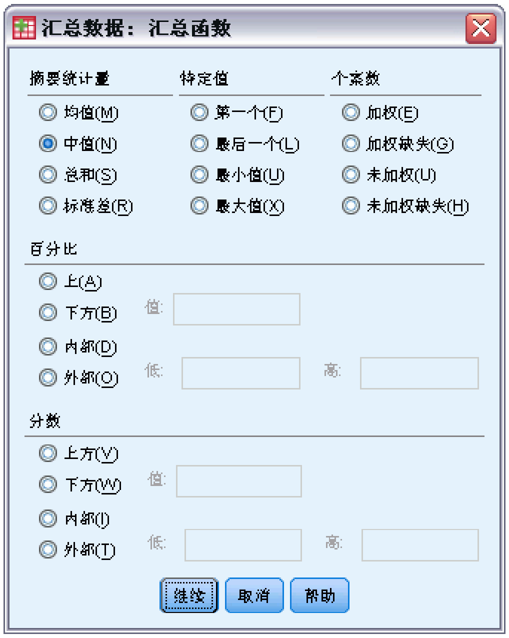
在汇总之前排序文件. 在使用大数据文件时（这种情况很少见），您可能发现有必要在进行汇总之前按照中断变量的值对数据文件进行排序。建议不要使用该选项，除非您遇到了内存或性能问题。

分类汇总数据：分类汇总函数

此对话框指定的函数用于计算“分类汇总数据”对话框中的“分类汇总变量”列表中选定变量的分类汇总数据值。分类汇总函数包括：

- 数值变量的摘要函数，包括均值、中位数、标准差和总和
- 个案数，包括未加权、加权、非缺失和缺失
- 在指定值以上或以下的值的百分比或分数
- 在指定范围之内或之外的值的百分比或分数

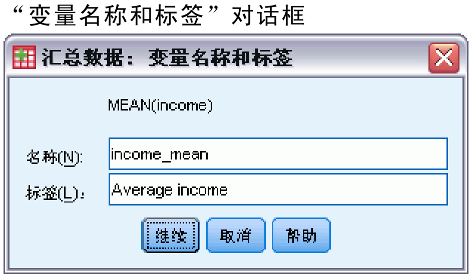
图片 9-8
“分类汇总函数”对话框



分类汇总数据：变量名称和标签

“分类汇总数据”指定新的数据文件中分类汇总变量的缺省变量名称。此对话框使您能够更改“分类汇总变量”列表上选定变量的名称并提供描述性变量标签。

图片 9-9
“变量名称和标签”对话框



分割文件

“拆分文件”将数据文件分割为单独的组，以根据一个或多个分组变量的值进行分析。如果您选择了多个分组变量，个案按“分组依据”列表的前一个变量的类别中的每个变量进行分组。例如：如果您选择性别作为第一个分组变量，选择少数民族作为第二个分组变量，将按每个性别类别中的少数民族分类对个案进行分组。

- 您最多可指定 8 个分组变量。

- 长字符串变量（长于 8 个字节的字符串变量）的每 8 个字节计为一个变量，直到达到 8 个分组变量的限制。
- 个案应按分组变量的值并以“分组依据”列表中列出变量的相同顺序排序。如果数据文件未准备好进行排序，请选择**按分组变量排序文件**。

图片 9-10
“拆分文件”对话框



比较各组。将拆分文件组显示在一起以用于比较目的。对于枢轴表，将创建单个数据枢轴表，且可以将每个拆分文件变量在表的维度之间移动。对于图表，为每个拆分文件组分别创建图表，并在“浏览器”中将图表显示在一起。

按组组织输出。为每个拆分文件组分别显示每个过程中的所有结果。

要拆分数据文件以进行分析，请执行下列操作

▶ 从菜单中选择：
数据 > 拆分文件...

▶ 选择**比较各组**或**按组来组织输出**。

▶ 选择一个或多个分组变量。

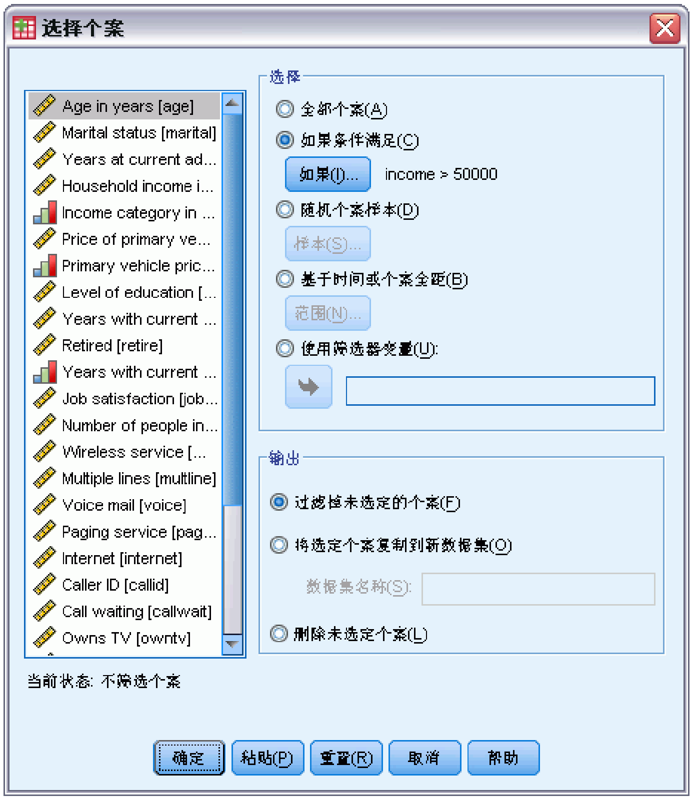
选择个案

“选择个案”提供了几种基于包含变量和复杂表达式的条件来选择个案子组的方法。您也可以选择个案的随机样本。用于定义子组的条件可包括：

- 变量值和范围
- 日期和时间范围
- 个案（行）编号
- 算术表达式

- 逻辑表达式
- 函数

图片 9-11
“选择个案”对话框



全部个案. 关闭个案过滤，并使用全部个案。

如果条件满足. 使用条件表达式选择个案。如果条件表达式的结果为 true，则选择该个案。如果结果为 false 或 missing，则不选择该个案。

随机个案样本. 根据近似的百分比或精确的个案数来选择随机样本。

基于时间或个案全距. 根据个案数目范围或日期/时间范围来选择个案。

使用筛选器变量. 使用来自数据文件的选定数值变量作为过滤变量。选择过滤变量值不为 0 或 missing 的个案。

输出

此部分控制对未选定个案的处理。可以选择以下可选项之一来处理未选定个案：

- **过滤掉未选定的个案。**未选定的个案不包括在分析中，但保留在数据集中。如果关闭过滤，则在会话中稍后可以使用未选定个案。如果选择随机样本，或者基于条件表达式选择个案，则此项生成名为 filter_$ 的变量，对选定个案该变量的值为 1，对未选定个案该变量的值为 0。

- **将选定的个案复制到新数据集。**选定的个案复制到新数据集，原始数据集未受影响。未选中个案不包括在新数据集中，而在初始数据集中保持其初始状态。
- **删除未选定个案。**从数据集删除未选定个案。只有退出文件而不保存任何更改，然后重新打开文件，才能恢复删除的个案。如果保存对数据文件的更改，则会永久删除个案。

注意：如果删除未选定个案并保存文件，则无法恢复个案。

要选择个案的子集，请执行下列操作

▶ 从菜单中选择：

数据 > 选择个案...

▶ 选择一个选择个案的方法。

▶ 指定选择个案的条件。

选择个案:如果

此对话框允许您使用条件表达式选择个案的子集。条件表达式对每个个案都返回一个值：true、false 或 missing。

图片 9-12
“选择个案：If”对话框



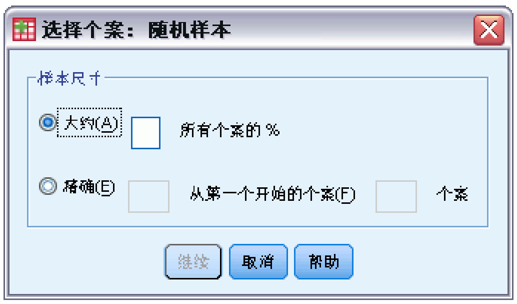
- 如果条件表达式的结果为 true，则所选子集中将包含该个案。
- 如果条件表达式结果为 false 或 missing，则所选子集中不包含该个案。

- 大多数条件表达式使用计算器面板上六个关系运算符（<、>、<=、>=、= 和 ~=）中的一个或多个。
- 条件表达式可以包含变量名称、常数、算术运算符、数值（和其他）函数、逻辑变量以及关系运算符。

选择个案:随机样本

此对话框允许您基于近似的百分比或准确的个案数来选择随机样本。抽样是在不进行替换的情况下执行的；所以同一个案不会被选择多次。

图片 9-13
“选择个案：随机样本”对话框



近似。生成近似于指定个案百分比的随机样本。由于此例程为每个个案作出独立的伪随机决策，因此选定个案的百分比只能近似于指定的百分比。数据文件中的个案越多，选定个案的百分比与指定百分比就越接近。

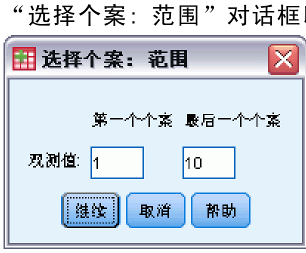
精确。用户指定的个案数。您还必须指定用于生成样本的个案数。第 2 个数目应该小于或等于数据文件中个案的总数。如果该数字超过数据文件中的个案总数，则样本所包含的个案数将按请求的数目成比例减少。

选择个案:范围

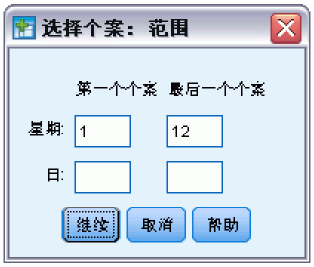
此对话框根据个案数目范围或日期/时间范围来选择个案。

- 个案范围基于在数据编辑器中显示的行号。
- 日期和时间范围仅可用于定义有日期变量（“日期”菜单，“定义日期”）的时间序列数据。

图片 9-14
“选择个案：范围”对话框以输入个案的范围（未定义 日期变量）



图片 9-15
“选择个案：范围”对话框以输入具有 已定义的日期变量的时间序列数据



加权个案

“对个案加权”给个案加以不同的权重（通过模拟复制）以进行统计分析。

- 加权变量的值应表明在您的数据文件中由单个个案代表的观察数据的个数。
- 加权变量值为 0、负值或缺失值的个案排除在分析之外。
- 小数值有效，在一些过程如“频率”、“交叉表”和“定制表”中会使用小数权重值。然而，大多数过程将权重变量视为重复权重，并将小数权重简单地四舍五入为最接近的整数。某些过程完全忽略权重变量，这一限制在特定过程文档中加以说明。

图片 9-16
“加权个案”对话框



一旦应用了一个权重变量，该权重变量始终保持有效，直到您选择另一个权重变量或关闭加权。如果您保存加权后的数据文件，加权信息随数据文件一起保存。您可随时关闭加权，即使在文件以加权形式保存之后也可以。

交叉表中的权重。“交叉表”过程有多个处理个案权重的选项。

散点图和直方图中的权重。散点图和直方图具有打开和关闭个案权重的选项，但对权重变量值为 0、负值或缺失值的个案没有影响。即使您在图表中关闭加权，这些个案仍将被排除在图表之外。

要对个案加权，请执行下列操作

▶ 从菜单中选择：

数据 > 加权个案...

▶ 选择**加权个案依据**。

▶ 选择一个频率变量。

该频率变量的值用作个案权重。例如，频率变量值为 3 的个案将在加权数据文件中代表 3 个个案。

重组数据

使用重组数据向导可为要使用的过程重组数据。该向导会使用新的重组过的文件替代当前文件。该向导可以：

- 将选定变量重组为个案(C)
- 将选定个案重组为变量(V)
- 转置所有数据(T)

重组数据

▶ 从菜单中选择：
数据 > 重组(R)...

▶ 选择要执行的重组的类型。

▶ 选择要重组的数据。

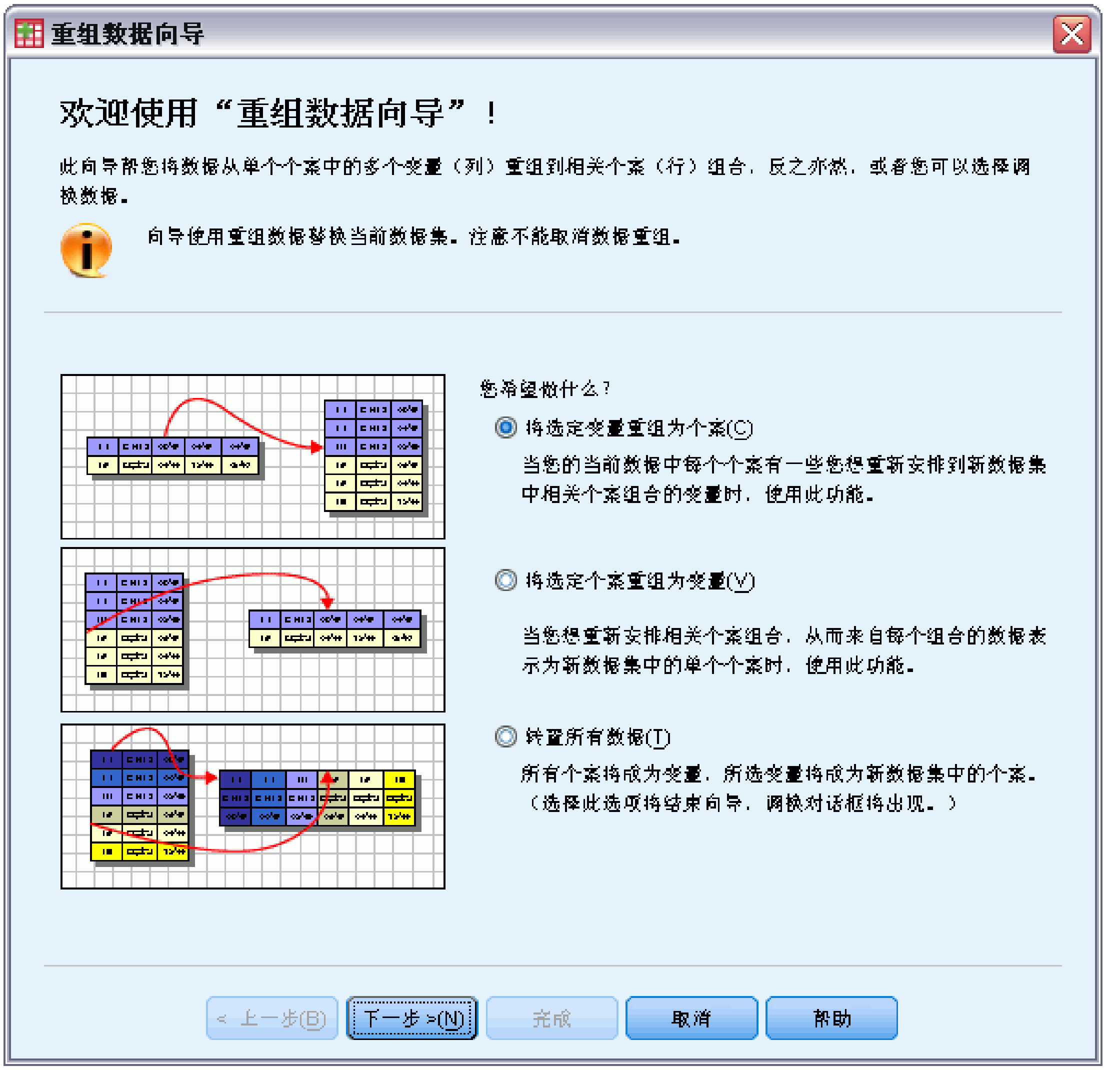
根据需要，您可以：

- 创建标识变量，这些变量允许您将新文件中的值追溯到原始文件中的值。
- 在重组之前对数据排序
- 为新文件定义选项
- 将命令语法粘贴到语法窗口中

重组数据向导：选择类型

使用重组数据向导可重组数据。在第一个对话框中，选择要执行的重组的类型。

图片 9-17
重组数据向导



- **将选定变量重组为个案。**如果数据中是相关列构成的组，并且希望其在新数据文件中显示为行组，可选择此项。如果选择此项，向导将显示**变量到个案**的步骤。
- **将选定个案重组为变量。**如果数据中是相关行构成的组，并且希望其在新数据文件中显示为列组，可选择此项。如果选择此项，向导将显示**个案到变量**的步骤。
- **转置所有数据。**要转置数据时选择此项。在新数据中，所有行将成为列，而所有列将成为行。此选项会关闭重组数据向导并打开“转置数据”对话框。

决定如何重组数据

变量包含要分析的信息，例如度量或得分。**个案**是观察值，例如个体。在简单数据结构中，每个变量是数据中的单独一列，每个个案是单独一行。因此，举个例子，如果度量班里所有学生的考试得分，所有得分将仅显示在一列中，而对于每个学生均有一行。

分析数据时，您常常根据某个条件分析变量如何变化。该条件可以是特定的试验处理、人口统计、时间点或者其他。在数据分析中，感兴趣的条件常常称为**因子**。分析因子时，数据结构是复杂的。在数据中，关于变量的信息可能位于多个列中（例如，对于每个因子水平均有一列），或者，关于个案的信息可能位于多个行中（例如，对于每个因子水平均有一行）。重组数据向导可帮助您重组具有复杂数据结构的文件。

当前文件的结构以及要在新文件中使用的结构决定了您在向导中选择的选项。

数据在当前文件中是如何排列的？当前数据的排列可能是因子记录在单独的变量中（在个案组中），也可能是因子与变量记录在一起（在变量组中）。

- **个案组。**当前文件的变量和条件记录在单独的列中吗？例如：

变量	因子
8	1
9	1
3	2
1	2

在此示例中，开头两行是**个案组**，因为它们相关。它们包含相同因子水平的数据。在 IBM® SPSS® Statistics 数据分析中，当数据以这种方式构造时，因子常常称为**分组变量**。

- **列组。**当前文件的变量和条件记录在同一列中吗？例如：

var_1	var_2
8	3
9	1

在此示例中，这两列是**变量组**，因为它们相关。它们包含同一变量的数据 — var_1 对应于因子水平 1 而 var_2 对应于因子水平 2。在 SPSS Statistics 数据分析中，当数据以此方式构造时，因子常常称为**重复度量**。

数据在新文件中应如何排列？这通常由要用来分析数据的过程决定。

- **要求使用个案组的过程。**要执行需要分组变量的分析，数据必须按个案组构造。示例包括“一般线性模型”的单变量、多变量和方差成分；“混合模型”；“OLAP 立方”；以及“T 检验”或“非参数检验”的独立样本。如果当前数据结构是变量组，要执行这些分析，请选择**将选定变量重组为个案**。
- **要求使用变量组的过程。**要分析重复度量，数据必须按变量组构造。示例包括“一般线性模型”的重复度量、“Cox 回归分析”的依时协变量 分析、“T 检验”的配对样本 或“非参数检验”的相关样本。如果当前数据结构是个案组，要执行这些分析，请选择**将选定个案重组为变量**。

变量到个案的示例

在此示例中，每个因子的检验得分记录在单独的列（A 和 B）中。

图片 9-18
变量到个案的当前数据

	score_a	score_b
1	1014.00	864.00
2	684.00	636.00
3	810.00	638.00

您想要执行独立样本 T 检验。您有一个由 score_a 和 score_b 组成的列组，但没有该过程需要的**分组变量**。在重组数据向导中选择**将选定变量重组为个案**，将一个变量组重组为名为 score 的变量，并创建名为 group 的索引。新的数据文件显示在下图中。

图片 9-19
变量到个案的新的、重组的数据

	组	得分
1	SCORE_A	1014.00
2	SCORE_B	864.00
3	SCORE_A	684.00
4	SCORE_B	636.00
5	SCORE_A	810.00
6	SCORE_B	638.00

如果运行独立样本 T 检验，则现在可以使用 group 作为分组变量。

个案到变量的示例

在此示例中，为每个主体记录两次检验得分，即处理之前和之后。

图片 9-20
个案到变量的当前数据

	主题	得分	时间
1	1	1014.00	bef
2	1	864.00	aft
3	2	684.00	bef
4	2	636.00	aft

您想要执行配对样本 T 检验。数据结构是个案组，但没有过程需要的成对变量的**重复度量**。在重组数据向导中选择**将选定个案重组为变量**，使用 id 标识当前数据中的组，并使用 time 创建新文件中的变量组。

图片 9-21
个案到变量的新的、重组的数据

	ID	之前	之后
1	1	864.00	1014.00
2	2	636.00	684.00

如果运行配对样本 T 检验，则现在可将 bef 和 aft 用作变量对。

重组数据向导（变量到个案）：变量组的数目

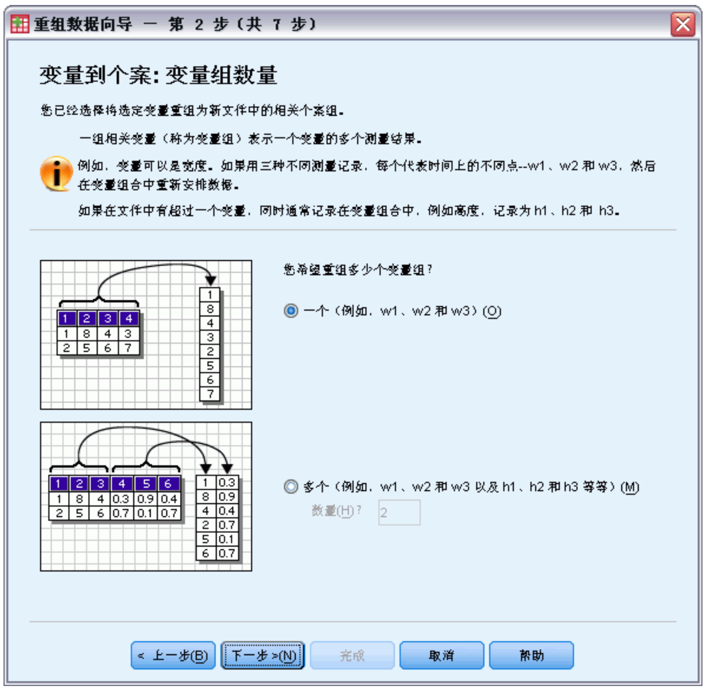
注意：如果选择将变量组重组为行，则向导会提供此步骤。

在此步骤中，决定要将当前文件中的多少变量组重组到新文件中。

当前文件中有多少个变量组？考虑有多少变量组存在于当前数据中。相关列组（称为**变量组**）在单独的列中记录同一个变量的重复度量。例如，如果当前数据中有三列（w1、w2 和 w3）记录**宽度**，则有一个变量组。如果有另外三列（h1、h2 和 h3）记录**高度**，则有两个变量组。

新文件中应有多少个变量组？考虑您想要让新数据文件中出现多少个变量组。您无需将所有变量组重组到新文件中。

图片 9-22
重组数据向导：变量组的数目，第 2 步



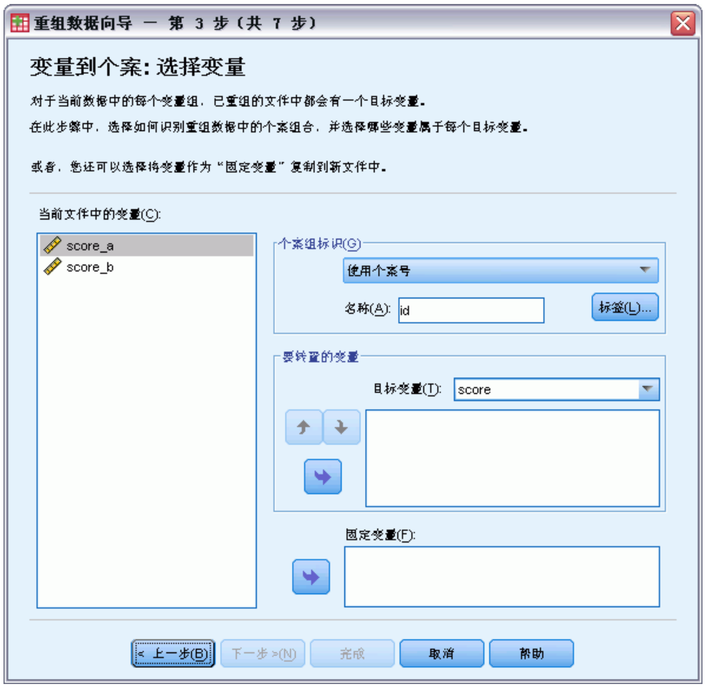
- **一个**。向导将从当前文件中的一个变量组在新文件中创建单个重组的变量。
- **多个**。向导将在新文件中创建多个重组的变量。所指定的个数会影响下一步，在下一步中向导将自动创建指定个数的新变量。

重组数据向导（变量到个案）：选择变量

注意：如果选择将变量组重组为行，则向导会提供此步骤。

在此步骤中，提供关于应如何在新文件中使用当前文件中的变量的信息。您还可以创建在新文件中标识行的变量。

图片 9-23
重组数据向导：选择变量，第 3 步



应如何标识新行？您可以在新数据文件中创建一个变量，该变量标识当前数据文件中用来创建一组新行的行。该标识符可以是连续的个案号，也可以是变量的值。使用“个案组标识”中的控件可定义新文件中的标识变量。单击单元格可更改默认变量名并为标识变量提供描述性变量标签。

新文件中应重组哪些内容？在前一步中，您告诉了向导要重组多少个变量组。向导已为每个组分别创建了一个新变量。变量组的值将出现在新文件的该变量中。使用“要转置的变量”中的控件可定义新文件中的重组的变量。

指定一个重组的变量

▶ 将组成要转换的变量组的变量放到“要转置的变量”列表中。组中的所有变量必须是同一类型（数值或字符串）。

在变量组中可以将同一个变量包含多次（变量是从源变量列表复制而非移动），该变量的值会在新文件中重复。

指定多个重组的变量

▶ 从“目标变量”下拉列表中选择要定义的第一个目标变量。

▶ 将组成要转换的变量组的变量放到“要转置的变量”列表中。组中的所有变量必须是同一类型（数值或字符串）。可以在变量组中将同一个变量包含多次。（变量是从源变量列表复制而非移动，其值在新文件中重复。）

▶ 选择下一个要定义的目标变量，并为提供的所有目标变量重复变量选择过程。

- 尽管可以在同一个目标变量组中将同一个变量包含多次，但不能将同一个变量包含在多个目标变量组中。
- 每个目标变量组列表必须包含相同个数的变量。（多次列出的变量包含在计数中。）
- 目标变量组的个数由您在前一步中指定的变量组的个数决定。在此处您可以更改默认变量名，但必须返回到前一步才能更改要重组的变量组的个数。
- 必须为提供的所有目标变量定义变量组（方法是在源列表中选择变量），然后才能继续下一步。

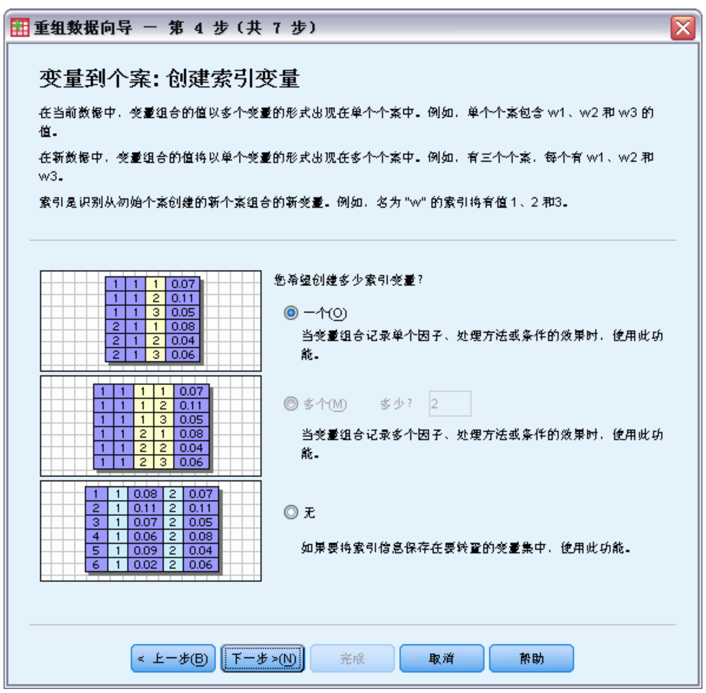
应将哪些内容复制到新文件中？未重组的变量可以被复制到新文件中。它们的值将复制到新的行。将要复制到新文件中的变量移至“固定变量”列表中。

重组数据向导（变量到个案）：创建索引变量

注意：如果选择将变量组重组为行，则向导会提供此步骤。

在此步骤中，决定是否创建索引变量。索引是一个新变量，它根据从中创建新行的原始变量按顺序标识行组。

图片 9-24
重组数据向导：创建索引变量，第 4 步



新文件中应有多少个索引变量？索引变量可以用作过程中的分组变量。在大多数情况下，单个索引变量就已足够，但是，如果当前文件中的变量组反映多个因子水平，则使用多个索引可能比较合适。

- **一个。**向导将创建单个索引变量。

- **多个。**向导将创建多个索引，请输入要创建的索引数。所指定的个数会影响下一步，在下一步中向导将自动创建指定个数的索引。
- **无。**如果不想在新文件中创建索引变量，则选择此项。

变量到个案的单索引示例

在当前数据中，有一个变量组宽度和一个因子时间。度量宽度三次，并记录到 w1、w2 和 w3 中。

图片 9-25
单索引的当前数据

	主题	w1	w2	w3
1	1	6.70	4.30	5.70
2	2	7.10	5.90	5.60

我们会将变量组重组为单个变量宽度，并创建单个数字索引。新的数据显示在下表中。

图片 9-26
带有一个索引的新的重组的数据

	主题	指标	宽度
1	1	1	6.70
2	1	2	4.30
3	1	3	5.70
4	2	1	7.10
5	2	2	5.90
6	2	3	5.60

索引从 1 开始，对组中的每个变量递增。每次遇到原始文件中的新行时，它会重新开始。现在就能在需要分组变量的过程中使用索引了。

变量到个案的双索引示例

当变量组记录多个因子时，您可以创建多个索引；但是，当前数据的排列必须使第一个因子的水平为主索引，后续因子的水平在主索引内循环。在当前数据中，有一个变量组宽度和两个因子 A 和 B。对数据的排列使因子 B 的水平在因子 A 的水平内循环。

图片 9-27
双索引的当前数据

	主题	w_a1b1	w_a1b2	w_a2b1	w_a2b2
1	1	5.50	6.40	5.80	5.90
2	2	7.40	7.10	5.60	6.70

我们会将变量组重组到单个变量宽度，并创建两个索引。新的数据显示在下表中。

图片 9-28
带有两个索引的新的重组的数据

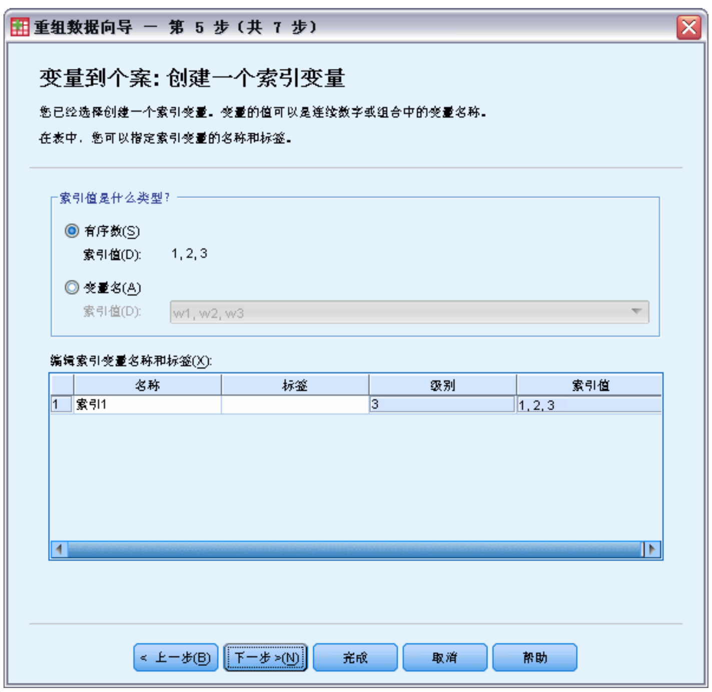
	主题	指标 a	指标 b	宽度
1	1	1	1	5.50
2	1	1	2	6.40
3	1	2	1	5.80
4	1	2	2	5.90
5	2	1	1	7.40
6	2	1	2	7.10
7	2	2	1	5.60
8	2	2	2	6.70

重组数据向导（变量到个案）：创建一个索引变量

注意：如果选择将变量组重组为行并创建一个索引变量，则向导会提供此步骤。

在此步骤中，决定要让索引变量取什么值。值可以是顺序编号或原始变量组中变量的名称。还可以为新的索引变量指定名称和标签。

图片 9-29
重组数据向导：创建一个索引变量，第 5 步



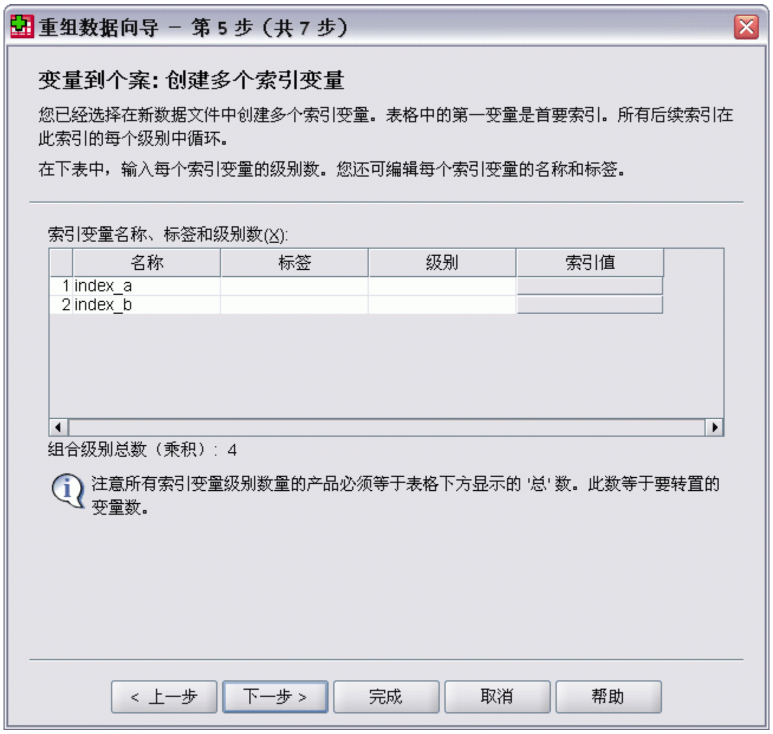
- **有序数。**向导会自动将顺序编号指定为索引值。
- **变量名称。**向导将使用选定变量组的名称作为索引值。请从列表中选择一个变量组。
- **名称和标签。**单击单元格可更改默认变量名并为索引变量提供描述性变量标签。

重组数据向导（变量到个案）：创建多个索引变量

注意：如果选择将变量组重组到行中并创建多个索引变量，则向导会提供此步骤。

在此步骤中，为每个索引变量指定水平数。还可以为新的索引变量指定名称和标签。

图片 9-30
重组数据向导：创建多个索引变量，第 5 步



当前文件中记录了多少个水平？考虑当前数据中记录了多少个因子水平。一个**水平**定义一组具有完全相同的条件的个案。如果有多个因子，则当前数据的排列必须使得第一个因子的水平是主索引，后续因子的水平在其内循环。

新文件中应有多少个水平？输入每个索引的水平数。多个索引变量的值始终是顺序编号。这些值从 1 开始，对每个水平递增。第一个索引递增最慢，最后一个索引递增最快。

总计组合水平。创建的水平数不能超过当前数据中存在的水平数。因为重组的数据中每个处理组合将包含一行，所以向导会检查您创建的水平数。它会将您创建的水平的乘积与变量组中的变量数进行比较。这两者必须匹配。

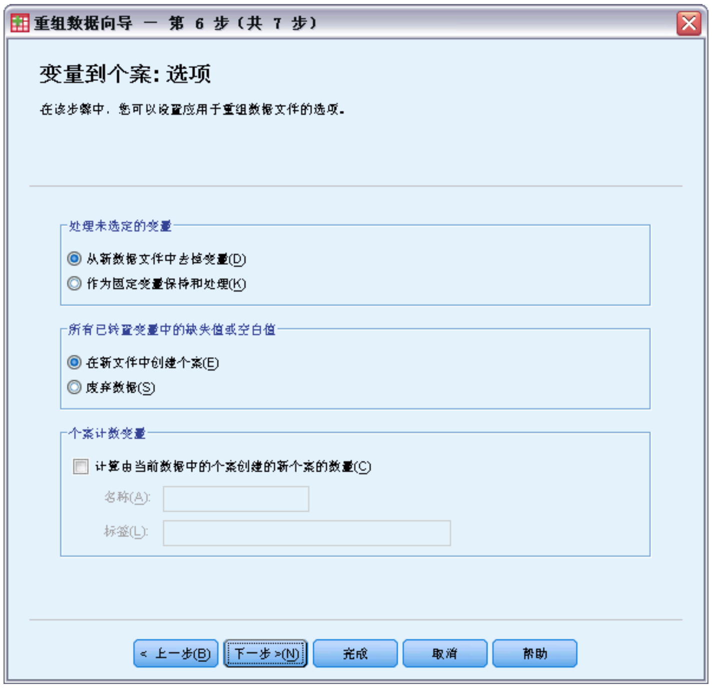
名称和标签。单击单元格可更改默认变量名并为索引变量提供描述性变量标签。

重组数据向导（变量到个案）：选项

注意：如果选择将变量组重组为行，则向导会提供此步骤。

在此步骤中，指定新的重组的文件的选项。

图片 9-31
重组数据向导：选项，第 6 步



删除未选中的变量？在选择变量的步骤中（第 3 步），从当前数据中选择了要重组的变量组、要复制的变量以及一个标识变量。选定变量中的值将出现在新文件中。如果当前数据中有其他变量，您可以选择是丢弃它们还是保留它们。

保留缺失值？向导在每个可能的新行中检查是否有空值。**空值**是系统缺失值或空白值。您可以选择是保留还是丢弃只包含空值的行。

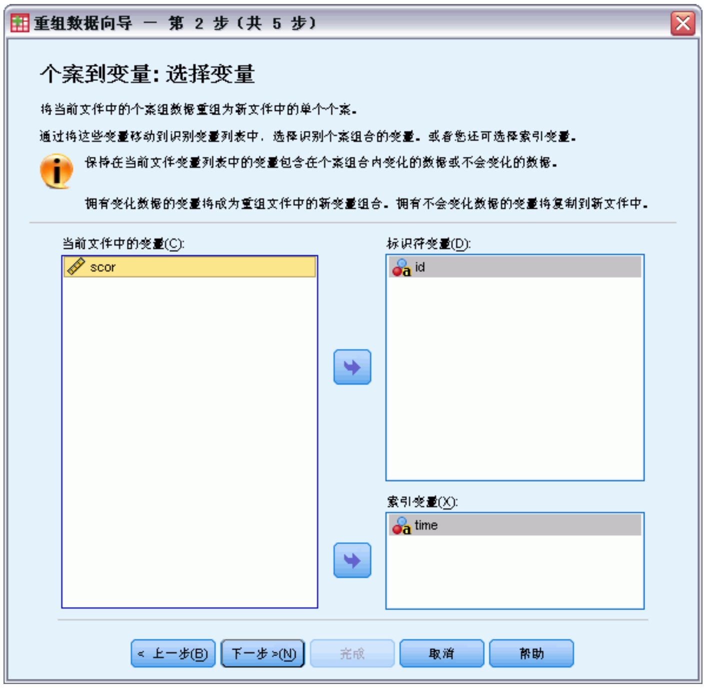
创建计数变量？向导可在新文件中创建一个**计数变量**。它包含由当前数据中的行所生成的新行的数目。如果选择丢弃新文件中的空值，则计数变量可能会有用，因为这样可对当前文件中的给定行生成不同数目的新行。单击单元格可更改默认变量名并为计数变量提供描述性变量标签。

重组数据向导（个案到变量）：选择变量

注意：如果选择将个案组重组为列，则向导会提供此步骤。

在此步骤中，提供关于应如何在新文件中使用当前文件中的变量的信息。

图片 9-32
重组数据向导：选择变量，第 2 步



什么标识当前数据中的个案组？一个个案组是一组行，这些行由于度量同一个观察单元（例如个体或机构）而相关。向导需要知道当前文件中哪些变量标识个案组，以便能将每个组合并到新文件的单独一行中。将在当前文件中标识个案组的变量移至“标识变量”列表中。用来分割当前数据文件的变量被自动用来标识个案组。每次遇到标识值的新组合时，向导将创建一个新行，因此当前文件中的个案应按标识变量的值排序，顺序与在“标识变量”列表中列出变量的顺序相同。如果当前文件尚未排序，可以在下一步对其进行排序。

应如何在新文件中创建新的变量组？在原始数据中，变量出现在单独一列中。在新的数据文件中，该变量将出现在多个新列中。**索引变量**是向导应用来创建新列的当前数据中的变量。重组的数据中将包含与这些列中的唯一值一一对应的新变量。将应该用来组成新变量组的变量移至“索引变量”列表中。当向导提供选项时，还可以选择按索引对新列排序。

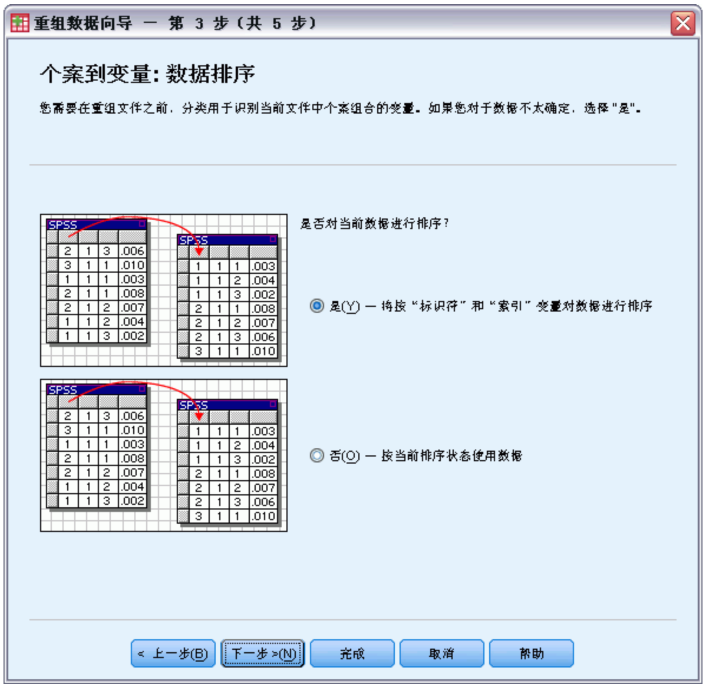
其他列会怎样？向导将自动决定如何处理保留在“当前文件”列表中的变量。它检查每个变量，查看数据值在个案组中是否变化。如果有变化，向导会将这些值重组到新文件中的一个变量组中。如果无变化，向导会将这些值复制到新文件。当确定某个变量在组中是否变化时，用户缺失值当作有效值处理，但系统缺失值则不然。如果该组包含一个有效或用户缺失值和系统缺失值，则会将其当作在组中无变化的变量来处理，向导会将这些值复制到新文件。

重组数据向导（个案到变量）：排序数据

注意：如果选择将个案组重组为列，则向导会提供此步骤。

在此步骤中，决定在重组当前文件之前是否对其排序。每次向导遇到新的标识值组合时，都会创建新行，因此按标识个案组的变量对数据排序就很重要。

图片 9-33
重组数据向导：对数据排序，第 3 步



行在当前文件中是如何排序的？考虑当前数据如何排序，以及使用哪些变量标识个案组（在前一步中指定）。

- **是**。向导将自动按标识变量对当前数据排序，其顺序与前一步中变量在"标识变量"列表中列出的顺序相同。当数据未按标识变量排序，或者您不确定时选择此项。此选项要求分开传递数据，但可保证对行正确排序以便重组。
- **否**。向导不会对当前数据排序。如果确信当前数据已按标识个案组的变量进行了排序，则选择此项。

重组数据向导（个案到变量）：选项

注意：如果选择将个案组重组为列，则向导会提供此步骤。

在此步骤中，指定新的重组的文件的选项。

图片 9-34
重组数据向导：选项，第 4 步



应如何在新文件中对新变量组进行排序？

- **按变量**。向导将从一个原始变量创建的新变量分组在一起。
- **按索引**。向导根据索引变量的值对变量分组。

示例。要重组的变量是 w 和 h，索引是月份：

w　　　h　　　月份

按变量分组的结果是：

w.jan　　　w.feb　　　h.jan

按索引分组的结果是：

w.jan　　　h.jan　　　w.feb

创建计数变量？向导可在新文件中创建一个计数变量。它包含用于在新数据文件中创建行的当前数据中的行数。

创建指示变量？向导可以使用索引变量在新的数据文件中创建**指示变量**。它为索引变量的每个唯一值创建一个新变量。指示变量指示个案存在值还是缺少值。如果个案有值，则指示变量的值为 1；否则为 0。

示例。索引变量为产品。它记录客户购买的产品。原始数据为：

客户	产品
1	小鸡
1	鸡蛋
2	鸡蛋
3	小鸡

创建指示变量会生成与产品的唯一值一一对应的新变量。重组的数据为：

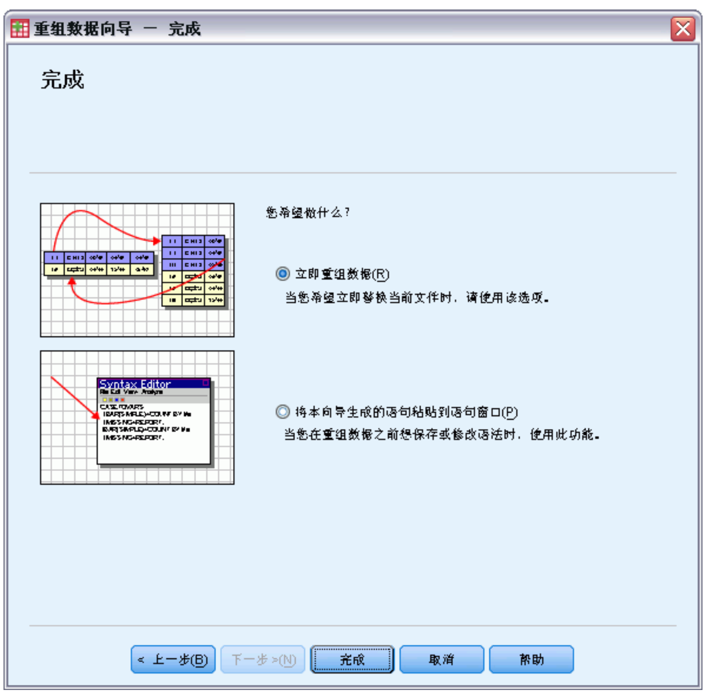
客户	indchick	indeggs
1	1	1
2	0	1
3	1	0

在此示例中，重组的数据可以用来获取客户所购买产品的频率计数。

重组数据向导：完成

这是重组数据向导的最后一步。决定如何处理您的指定。

图片 9-35
重组数据向导：完成



- **立即重组。**向导将创建新的重组的文件。如果想要立即替换当前文件，则选择此项。

注意：如果原始数据已加权，则除非已重组了用作权重的变量或者已将其从新文件删除，否则新数据也会加权。

- **粘贴语法**。向导会将其生成的语法粘贴到语法窗口。当尚未准备好替换当前文件时，想要修改语法时，或者想要保存语法以供将来使用时，选择此项。

章

10

使用输出

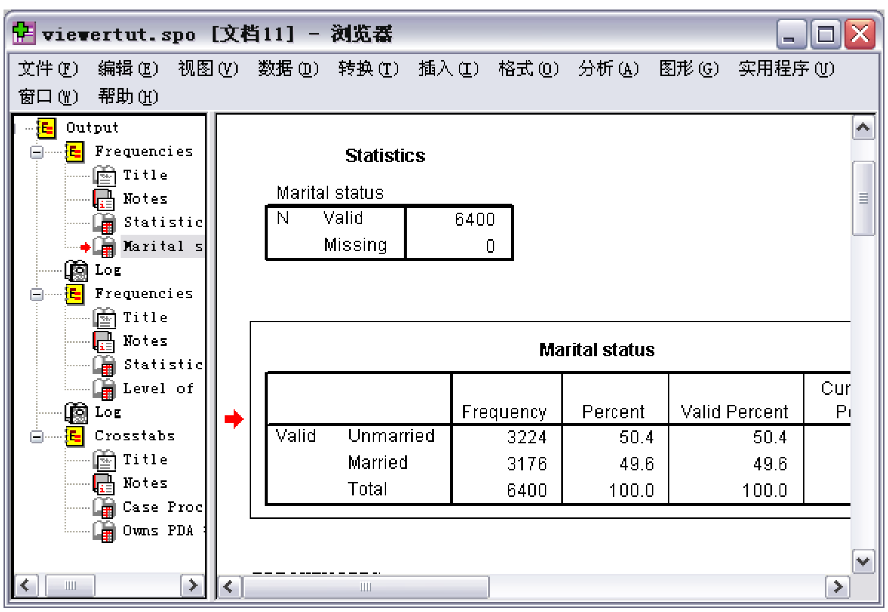
运行过程时，结果显示在称为“查看器”的窗口中。在该窗口中，可以很容易地浏览到您要查看的输出。您也可以使用输出并创建包含所需的确切输出内容的文档。

查看器

结果显示在查看器中。您可以将查看器用于：

- 浏览结果
- 显示或隐藏选中的表格和图表
- 通过移动选中的项更改结果的显示顺序
- 在查看器和其他应用程序间移动各项

图片 10-1
查看器



查看器分为两个窗格：

- 左窗格包含内容的概要视图。
- 右窗格包含统计表、图表和文本输出。

可以单击概要中的项直接转到相应的表格或图表。可以单击并拖动概要窗格的右边框来更改概要窗格的宽度。

显示和隐藏结果

在查看器中，可以选择显示或隐藏单个表格或来自整个过程的结果。此过程在要减少内容窗格中的可视输出的量时很有用。

隐藏表格和图表

▶ 双击查看器的概要窗格中项的书形图标。

或

▶ 单击某项将其选中。

▶ 从菜单中选择：
视图 > 隐藏

或

▶ 单击“概要”工具栏上的合上的书形（隐藏）图标。

打开的书形（显示）图标变为活动图标，表示该项现在已隐藏。

隐藏过程结果

▶ 在概要窗格中单击过程名称左边的框。

这将隐藏程序的所有结果，从而折叠概要视图。

移动、删除和复制输出

可以通过复制、移动或删除一项或一组项来重新排列结果。

在查看器中移动输出

▶ 选择概要窗格或内容窗格中的某项。

▶ 将选中的项拖放到不同的位置。

在查看器中删除输出

▶ 选择概要窗格或内容窗格中的某项。

▶ 按 Delete 键。

或

▶ 从菜单中选择：
编辑 > 删除

更改初始对齐方式

默认情况下，所有的结果最初都是左对齐的。要更改新输出项的初始对齐方式：

▶ 从菜单中选择：
编辑 > 选项

▶ 单击**查看器**选项卡。

▶ 在“初始输出状态”组中，选择项类型（例如，枢轴表、图表、文本输出）。

▶ 选择所需的对齐选项。

更改各输出项的对齐方式

▶ 在概要或内容窗格中，单击要对齐的项。

▶ 从菜单中选择：
格式 > 左对齐

或
格式 > 居中

或
格式 > 右对齐

查看器概要

概要窗格提供查看器文档的目录。可以使用概要窗格浏览结果并控制显示。概要窗格中的大多数操作对内容窗格有相应的影响。

- 在概要窗格中选择一项将在内容窗格中显示相应的项。
- 在概要窗格中移动一项将移动内容窗格中相应的项。
- 折叠概要视图将隐藏折叠级别中的所有项的结果。

控制概要显示。要控制概要显示，您可以：

- 展开和折叠概要视图
- 更改选定各项的概要级别
- 更改概要显示中的各项的大小
- 更改概要显示中使用的字体

折叠和展开概要视图

▶ 单击要折叠或展开的概要项左边的框。

或

▶ 单击概要中的项。

▶ 从菜单中选择：
视图 > 折叠

或
视图 > 展开

更改概要级别

- ▶ 在概要窗格中单击项。
- ▶ 单击“概要”工具栏上的左箭头，使该项的级别上升（将项向左移）。

  或

  单击“概要”工具栏上的右箭头，使该项的级别下降（将项向右移）。

  或
- ▶ 从菜单中选择：
  编辑 > 大纲 > 提升

  或
  编辑 > 大纲 > 降级

更改概要级别在您在概要级别中移动了各项后特别有用。移动各项会更改各项的概要级别，您可以使用“概要”工具栏上的左箭头和右箭头来恢复原始的概要级别。

更改概要项的大小

- ▶ 从菜单中选择：
  视图 > 概要尺寸
- ▶ 选择概要大小（小、中或大）。

更改概要中的字体

- ▶ 从菜单中选择：
  视图 > 概要字体...
- ▶ 选择一种字体。

将项添加到查看器

在查看器中，您可以添加项，如标题、新文本、图表或来自其他应用程序的材料。

添加标题或文本

可以将没有连接到表格或图表的文本项添加到查看器。

- ▶ 单击将放在标题或文本前面的表格、图表或其他对象。
- ▶ 从菜单中选择：
  插入 > 新建标题

  或
  插入 > 新建文本
- ▶ 双击该新对象。

▶ 输入文本。

添加文本文件

▶ 在查看器的概要窗格或内容窗格中，单击要放在文本前面的表格、图表或其他对象。

▶ 从菜单中选择：
插入 > 文本文件...

▶ 选择文本文件。

要编辑文本，请双击该文本。

将对象粘贴到查看器

来自其他应用程序的对象可以粘贴到查看器中。可以使用“之后粘贴”或“选择性粘贴”。两种类型的粘贴都把新对象放在查看器中当前选中的对象的后面。要选择粘贴对象的格式，请使用“选择性粘贴”。

查找和替换查看器中的信息

▶ 要在查看器中查找或替换信息，请从菜单中选择：
编辑 > 查找

或

编辑 > 替换

图片 10-2
“查找和替换”对话框



您可使用“查找和替换”来：

- 搜索整个文档或仅选中项。
- 从当前位置向下或向上搜索。
- 搜索窗格或限制搜索内容或概要窗格。
- 搜索隐藏项。其中包括内容窗格中隐藏的任何项（例如，默认情况下隐藏的“注释”表）和枢轴表中隐藏的行和列。
- 限制搜索标准至大小写匹配。
- 限制枢轴表中的搜索标准至整个单元格内容匹配。

隐藏项目和枢轴表层

- 多维枢轴表当前可视层以下的层并不视为隐藏，并且会包含在搜索范围中，即使隐藏项不包含在搜索中。
- 隐藏项目包括内容窗格中的隐藏项（概要窗格中有合上的书形图标的项，或包含在概要窗格的折叠块中的项）和枢轴表中默认情况下隐藏的行和列（例如，默认情况下隐藏的空行和空列）或手动隐藏的行和列（通过编辑表并选择隐藏特定行或列）。隐藏项目仅在您明确选择**包含隐藏项目**时包含在搜索中。
- 在两种情况下，包含搜索文本或值的隐藏或不可见元素会在找到时显示，但项目随后即返回其原始状态。

轻量表

“查找和更换”将在轻量表中查找指定值，但不会替换值，因为轻量表无法编辑。如果在“替换”操作期间从轻量表中找到指定值，则会发出警告提醒您在轻量表（不可编辑）中找到值但未能更换。有关轻量表的详细信息，请参阅 枢轴表选项

将输出复制到其他应用程序

输出对象可以复制并粘贴到其他应用程序中，如字处理程序或电子表格。可以粘贴多种格式的输出。根据目标应用程序，可以使用以下部分或所有的格式：

图片（元文件）。可以将枢轴表和图表作为元文件图片进行粘贴。可以在其他应用程序中调整图片格式的大小，有时还可以用其他应用程序中的工具对其进行有限的编辑。作为图片粘贴的枢轴表保持所有的边框和字体特征。对于枢轴表，从“选择性粘贴”菜单中选择**文件**。此格式只在 Windows 操作系统上可用。

RTF（丰富文本格式）。枢轴表可以用 RTF 格式粘贴到其他应用程序。在大多数应用程序中，这意味着将枢轴表作为表格粘贴，然后可在其他应用程序中编辑它。

注意：Microsoft Word 可能无法正确显示过宽的表格。

位图。图表以及其他图形可以作为位图粘贴到其他应用程序中。

BIFF。 表格的内容可以粘贴到电子表格中并保持数值精确度。此格式只在 Windows 操作系统上可用。

文本。表格的内容可以作为文本进行复制和粘贴。此过程对于电子邮件这样的应用程序很有用，因为这种应用程序只接受或传送文本。

如果目标应用程序支持多种可用格式，则可能包含允许您选择格式“选择性粘贴”菜单项，或自动显示可用格式列表。

将输出项复制并粘贴到其他应用程序

▶ 选择查看器的概要窗格或内容窗格中的对象。

▶ 从“查看器”菜单中选择：

编辑 > 粘贴

▶ 从目标应用程序的菜单中选择：

编辑 > 粘贴

或

编辑 > 选择性粘贴...

- **粘贴。**输出以多种格式复制到剪贴板。每个应用程序决定“粘贴”使用的“最佳”格式。
- **选择性粘贴。**结果用多种格式复制到剪贴板。“选择性粘贴”使您可以从目标应用程序可用的格式列表中选择您想要的格式。

复制和粘贴模型视图。 包括表格在内的所有模型视图粘贴为图形图像。

复制和粘贴枢轴表。当以 Word/RTF 格式粘贴枢轴表时，根据枢轴表选项设置，将对文档宽度过宽的表格进行换行、按比例缩小以适合文档宽度，或不进行更改。

导出输出

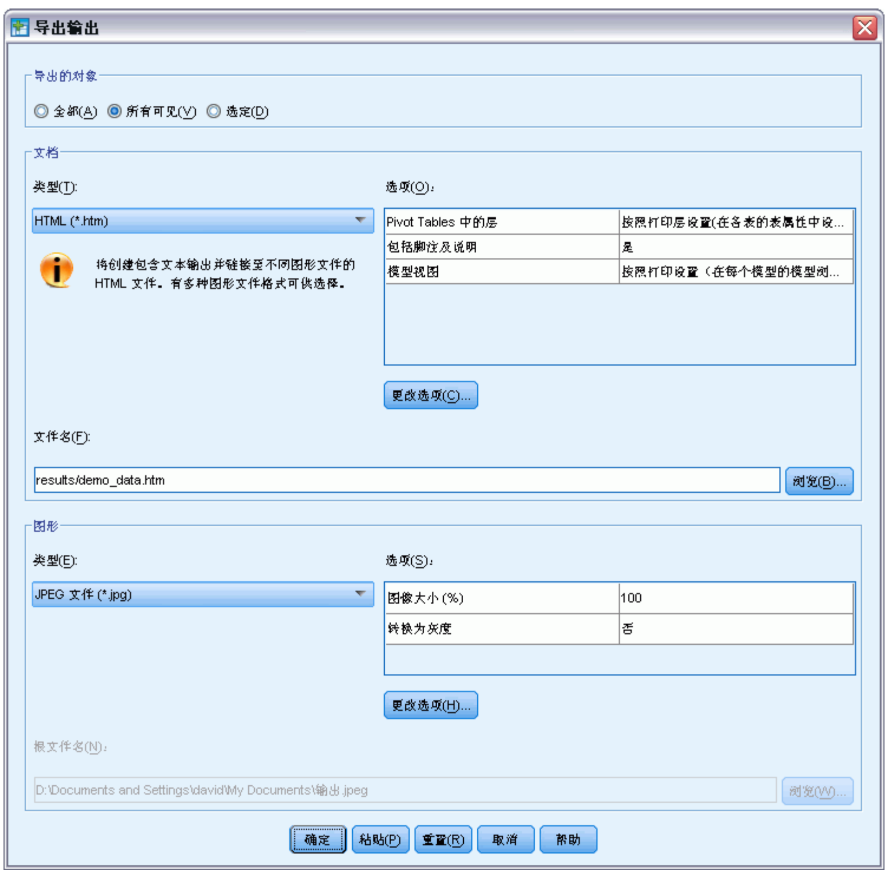
导出输出以 HTML、文本、Word/RTF、Excel、PowerPoint（需要 PowerPoint 97 或更高版本）和 PDF 格式保存查看器输出。图表也可以以多种不同图形格式导出。

注意：导出到 PowerPoint 功能仅在 Windows 操作系统中可用，且学生版未提供此功能。

导出输出

- ▶ 使查看器成为活动窗口（单击该窗口中的任意位置）。
- ▶ 从菜单中选择：
  文件 > 导出...
- ▶ 输入一个文件名（或图表前缀）并选择一种导出格式。

图片 10-3
“导出输出”对话框



要导出的对象。您可以导出查看器中的所有对象、所有可视对象或者仅选定的对象。

文档类型。可用选项包括：

- **Word/RTF (*.doc)**。枢轴表导出为 Word 表，并带有所有完整的格式属性，例如单元格边框、字体样式、背景色。文本输出导出为格式化的 RTF。图表、树形图和模型视图包括在 PNG 格式中。

  注意：Microsoft Word 可能无法正确显示过宽的表格。

- **Excel (*.xls)**。枢轴表行、列和单元格导出为 Excel 行、列和单元格，并带有所有完整的格式属性，例如单元格边框、字体样式、背景色等等。文本输出连同所有字体属性一起导出。文本输出中的一行即为 Excel 文件中的一行，文本输出中行的所有内容包含在单个单元格中。图表、树形图和模型视图包括在 PNG 格式中。
- **HTML (*.htm)**。枢轴表导出为 HTML 表。文本输出导出为预设置格式的 HTML。图表、树形图和模式视图按引用嵌入，您应以合适的格式导出图表，以包含在 HTML 文档中（例如，PNG 和 JPEG）。
- **可移植文件格式 (*.pdf)**。所有输出都将按“打印预览”中的显示导出，所有格式设置特性都不变。
- **PowerPoint 文件 (*.ppt)**。枢轴表导出为 Word 表，并嵌入到 PowerPoint 文件的各个幻灯片中，每个枢轴表对应一个幻灯片。将保留枢轴表的所有格式化属性，例如单元格边框、字体样式和背景色。图表、树形图和模型视图以 TIFF 格式导出。不包括文本输出。

  导出到 PowerPoint 仅在 Windows 操作系统上可用。

- **文本格式 (*.txt)**。文本输出格式包括纯文本、UTF-8 和 UTF-16。枢轴表可以以制表符分隔格式或空格分隔格式导出。所有文本输出以空格分隔格式导出。对于图表、树形图和模型视图，为每个图形将一行插入到文本文件中，以指示图像的文件名。
- **无（只有图形）**。 可用导出格式包括：EPS、JPEG、TIFF、PNG 和 BMP。在 Windows 操作系统上还可使用 EMF（增强型图元文件）格式。

输出管理系统。也可以将所有输出或用户指定的输出类型自动导出为 Word、Excel、PDF、HTML、文本或 IBM® SPSS® Statistics 格式的数据文件。

HTML 选项

以下选项可用于以 HTML 格式导出输出：

枢轴表中的层。缺省情况下，包含或排除枢轴表层由每个枢轴表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。

包括脚注及题注。控制包含或排除所有枢轴表脚注及题注。

模型视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。（注意：包括表在内的所有模型视图，均导出为图形。）

注意：对于 HTML，您还可以控制导出图表的图像文件格式。

设置 HTML 导出选项

▶ 选择 HTML 作为导出格式。

▶ 单击更改选项。

图片 10-4
HTML 导出输出选项



Word/RTF 选项

以下选项可用于以 Word 格式导出输出：

枢轴表中的层。缺省情况下，包含或排除枢轴表层由每个枢轴表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。

宽枢轴表。 控制当表相对于定义文档宽度时过宽的处理方法。缺省情况下，表将换行以适合宽度。表将划分为不同的部分，且行标签对每个表部分重复。或者，还可以缩小宽表，或不必改动宽表，而允许它们延伸到定义文档宽度以外。

包括脚注及题注。控制包含或排除所有枢轴表脚注及题注。

模型视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。（注意：包括表在内的所有模型视图，均导出为图形。）

导出页面设置。 这将打开一个对话框，在此可以定义导出文档的页面大小和边距。用于确定换行和缩小行为的文档宽度是页面宽度减去左右边距。

设置 Word 导出选项

▶ 选择 Word/RTF 作为导出格式。

▶ 单击更改选项。

图片 10-5
Word 导出输出选项



Excel 选项

以下选项可用于以 Excel 格式导出输出：

创建一个工作表或工作簿，或修改一个现有工作表。 默认情况下，将创建一个新工作簿。如果拥有指定名称的文件已存在，将被覆盖。如果您选择创建工作表的选项，且拥有指定名称的工作表已存在于指定文件中，则将被覆盖。如果您选择修改现有工作表的选项，您还必须指定工作表名称。（这对创建工作表为可选选项。）工作表名称不能超过 31 个字符，不能包含正斜杠或反斜杠、方括号、问号或星号。

如果您修改一个现有工作表，图表、模型视图和树形图不包括在导出输出中。

工作表中的位置。 控制导出输出在工作表中的位置。默认情况下，导出输出将添加到拥有任何内容的最后一列之后，从第一行开始，而不修改任何现有内容。这对于将新列添加到现有工作表是一个不错的选择。将导出输出添加到最后一行之后对于将新行添加到现有工作表是一个不错的选择。将导出输出添加到特定单元格位置的开始将覆盖添加导出输出的区域中的任何现有内容。

枢轴表中的层。缺省情况下，包含或排除枢轴表层由每个枢轴表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。

包括脚注及题注。控制包含或排除所有枢轴表脚注及题注。

模型视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。（注意：包括表在内的所有模型视图，均导出为图形。）

设置 Excel 导出选项

▶ 选择 Excel 作为导出格式。

▶ 单击更改选项。

图片 10-6
Excel 导出输出选项



PowerPoint 选项

对于 PowerPoint，可用选项有：

枢轴表中的层。缺省情况下，包含或排除枢轴表层由每个枢轴表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。

宽枢轴表。 控制当表相对于定义文档宽度时过宽的处理方法。缺省情况下，表将换行以适合宽度。表将划分为不同的部分，且行标签对每个表部分重复。或者，还可以缩小宽表，或不必改动宽表，而允许它们延伸到定义文档宽度以外。

包括脚注及题注。控制包含或排除所有枢轴表脚注及题注。

使用查看内略图项作为幻灯片标题。在导出操作创建的每个幻灯片上包括标题。每个幻灯片包含从“查看器”导出的单个项。标题是从查看器概要窗格中的项的概要条目构成的。

模型视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。（注意：包括表在内的所有模型视图，均导出为图形。）

导出页面设置。 这将打开一个对话框，在此可以定义导出文档的页面大小和边距。用于确定换行和缩小行为的文档宽度是页面宽度减去左右边距。

设置 PowerPoint 导出选项

▶ 选择 PowerPoint 作为导出格式。

▶ 单击更改选项。

图片 10-7
“Powerpoint 导出选项”对话框



注意：导出到 PowerPoint 仅在 Windows 操作系统上可用。

PDF 选项

对于 PDF，可用选项有：

内嵌书签。此选项提供 PDF 文档中对应于浏览器概要条目的书签。与浏览器概要窗格一样，通过书签可以更轻松地浏览具有大量输出对象的文档。

内嵌字体。内嵌字体确保 PDF 文档在所有计算机上外观相同。否则，如果文档中使用的一些字体在用于查看（或打印）PDF 文档的计算机上不可用，则替换字体的效果可能欠佳。

枢轴表中的层。缺省情况下，包含或排除枢轴表层由每个枢轴表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。

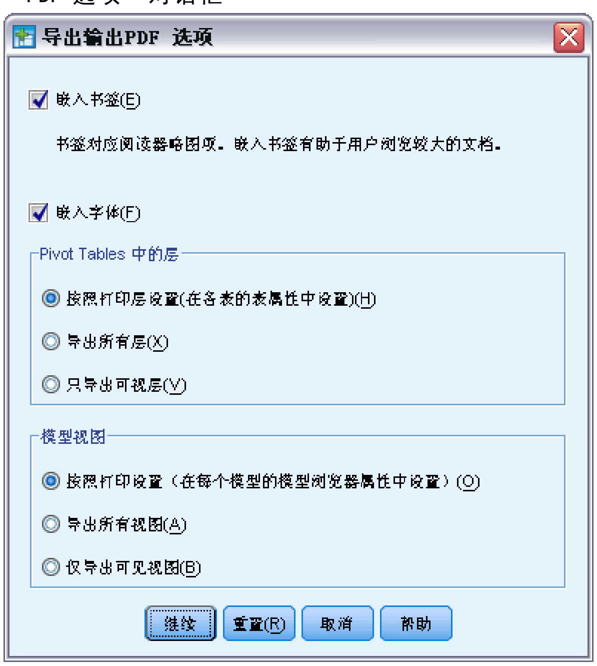
模型视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。（注意：包括表在内的所有模型视图，均导出为图形。）

设置 PDF 导出选项

▶ 选择**可移植文档格式**作为导出格式。

▶ 单击**更改选项**。

图片 10-8
“PDF 选项”对话框



其他影响 PDF 输出的设置

页面设置/页面属性。PDF 文档中的页面大小、方向、边距、页眉和页脚的内容和显示以及打印的图表大小都由“页面设置”和“页面属性”选项控制。

表格属性/表格外观。宽表和/或长表的比例调整和表层打印受每个表的表格属性控制。这些属性也可保存在表格外观中。

缺省/当前打印机。PDF 文档的分辨率 (DPI) 是缺省打印机或当前选定的打印机的当前分辨率设置（可以使用“页面设置”进行更改）。最高分辨率为 1200 DPI。如果打印机的设置更高，则 PDF 文档的分辨率将为 1200 DPI。

注意：在低分辨率的打印机上打印高分辨率文档时，输出结果的质量可能会比较差。

文本选项

对于文本输出，可用选项有：

枢轴表格式。枢轴表可以以制表符分隔格式或空格分隔格式导出。对于空格分隔格式，您还可以控制：

- **列宽。**自动适应不会对任何列内容换行，每列宽度为该列中最宽标签或值的宽度。定制为表格中所有列设置最大列宽，超过该宽度的值将折行显示在该列的下一行上。
- **行/列边框字符。**控制用于创建行和列边框的字符。如果不显示行和列边框，请将值留空。

枢轴表中的层。缺省情况下，包含或排除枢轴表层由每个枢轴表的表属性控制。您可以覆盖此设置，包含所有层或排除当前可视层之外的所有层。

包括脚注及题注。控制包含或排除所有枢轴表脚注及题注。

模型视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。（注意：包括表在内的所有模型视图，均导出为图形。）

设置文本导出选项

▶ 选择文本作为导出格式。

▶ 单击更改选项。

图片 10-9
“文本选项”对话框

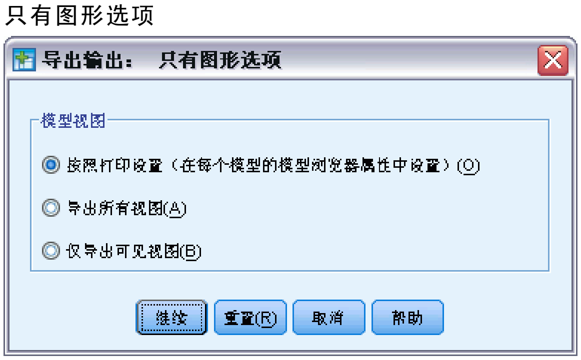


只有图形选项

对于只导出图形，可用选项有：

模型视图。 缺省情况下，包含或排除模型视图由每个模型的模型属性控制。您可以覆盖此设置，包含所有视图或排除当前可见视图之外的所有视图。（注意：包括表在内的所有模型视图，均导出为图形。）

图片 10-10
只有图形选项



图形格式选项

对于 HTML 与文本文档，如果仅导出图表，您可以选择图形格式，并为每种图形格式控制不同可选设置。

要选择导出图表的图形格式和选项：

- 选择 **HTML**、**文本**或**无（只有图形）**作为文档类型。
- 从下拉列表中选择图形文件格式。
- 单击**更改选项**更改选定图形文件格式选项。

JPEG 图表导出选项

- **图像大小。**原始图表大小的百分比，可达 200％。
- **转换为灰度。**将颜色转换成灰色阴影。

BMP 图表导出选项

- **图像大小。**原始图表大小的百分比，可达 200％。
- **压缩图像，以减小文件大小。**创建较小的文件而不影响图像质量的无损耗压缩技术。

PNG 图表导出选项

图像大小。原始图表大小的百分比，可达 200％。

颜色深度。决定导出图表中的颜色的数量。在任何深度下保存的图表最少的颜色数是实际使用的颜色数，最大则是深度允许的颜色数。例如，如果图表包含三种颜色：红色、白色和黑色，而您把它保存为 16 色，则图表仍将保持三种颜色。

- 如果图表中的颜色数超出了该深度的颜色数，则颜色将进行仿色处理，以复制图表中的颜色。
- **当前屏幕深度**是当前显示在计算机显示器上的颜色数。

EMF 和 TIFF 图表导出选项

图像大小。原始图表大小的百分比，可达 200％。

注意：EMF（增强型图元文件）格式只在 Windows 操作系统上可用。

EPS 图表导出选项

图像大小。您可以将大小指定为原图像大小的百分比（可达 200％），或指定图像的像素宽度（高度取决于宽值和宽高比）。导出图像总是与原始图像成比例。

包括 TIFF 预览图像。将包含 EPS 图像的预览保存为 TIFF 格式，便于在不能在屏幕上显示 EPS 图像的应用程序中进行显示。

字体。控制 EPS 图像中字体的处理。

- **使用字体参考。**如果图表中使用的字体在输出设备上可用，则使用这些字体。否则，输出设备将使用备用字体。
- **将字体替换为曲线。**将字体转换为 PostScript 曲线数据。文本本身已经不再像可编辑 EPS 图形的应用程序中的文本那样可编辑了。如果图表中使用的字体对输出设备不可用，则该选项十分有用。

查看器打印

有两个选项可用于打印“查看器”窗口的内容：

所有可见输出。仅打印当前显示在内容窗格中的项。不打印隐藏的项（概要窗格中有合上的书形图标的项，或者折叠的概要层中隐藏的项）。

选择。仅打印当前在概要窗格和/或内容窗格中选择的项。

打印输出和图表

▶ 使查看器成为活动窗口（单击该窗口中的任意位置）。

▶ 从菜单中选择：
文件 > 打印...

▶ 选择所需的打印设置。

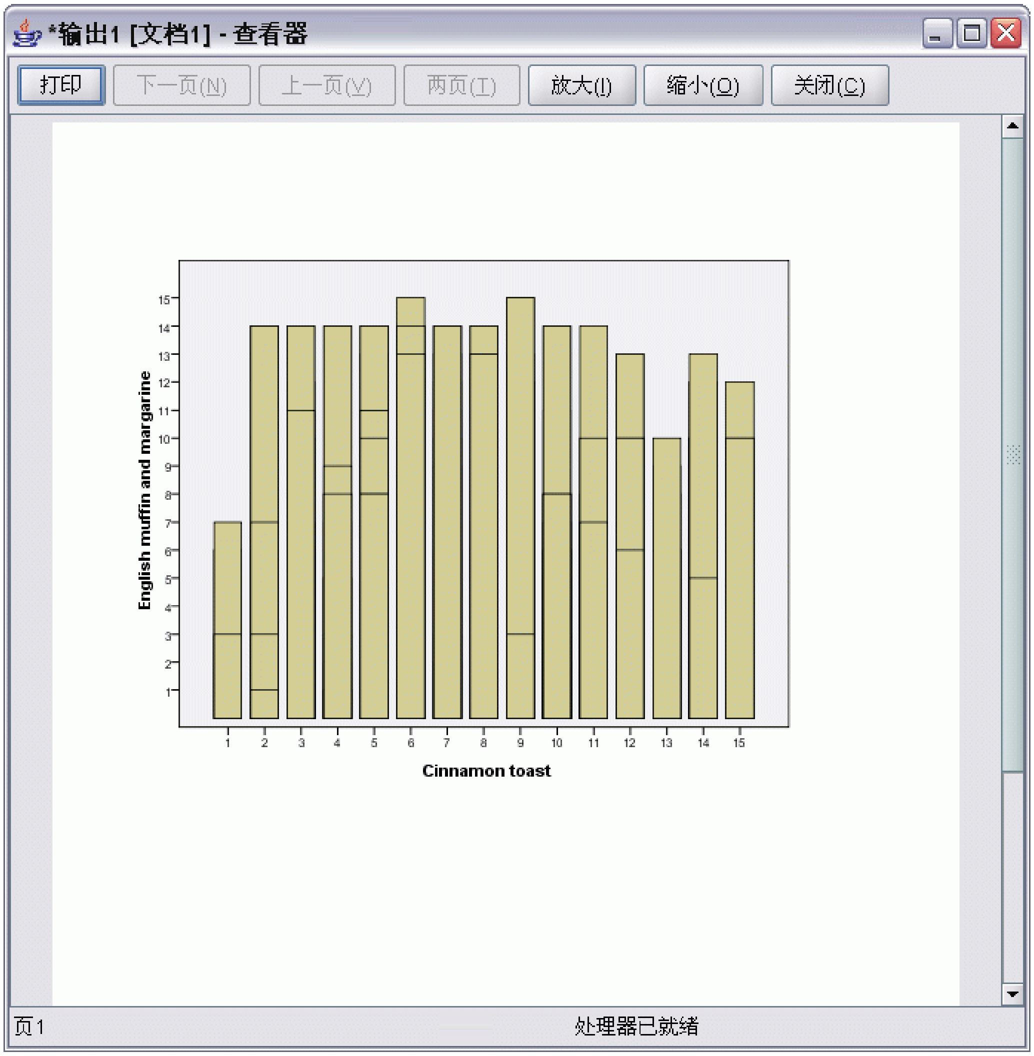
▶ 单击**确定**进行打印。

打印预览

“打印预览”显示查看器文档的每一页上将打印什么内容。我们建议您在实际打印查看器文档前查看“打印预览”，因为“打印预览”会显示通过查看查看器的内容窗格不可视的项，包括：

- 分页符
- 枢轴表隐藏的层
- 宽表中的分隔符
- 打印在每一页上的页眉和页脚

图片 10-11
打印预览



如果查看器中当前选择了任何输出，则预览仅显示选中的输出。要查看所有输出的预览，请确保查看器中没有选择任何输出。

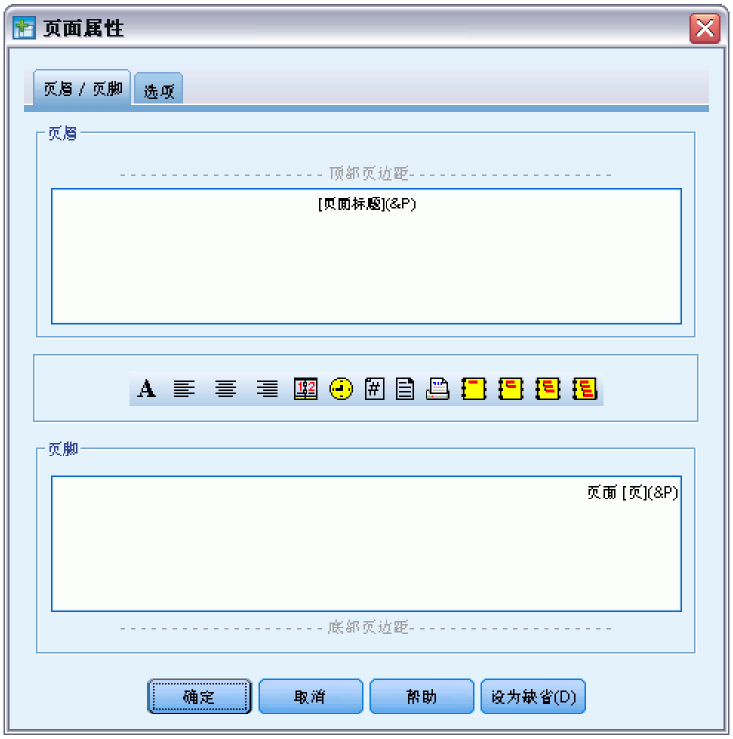
页面属性：页眉和页脚

页眉和页脚是打印在每一页的顶部和底部的信息。您可以输入任何文本用作页眉和页脚。还可以使用对话框中间的工具栏插入：

- 日期和时间
- 页码
- 查看器文件名
- 概要标题标签

- 页面标题和子标题

图片 10-12
“页面属性”对话框，“页眉/页脚”选项卡



- **设为默认**使用此处指定的设置作为新查看器文档的默认设置。（注意：这使页眉/页脚选项卡和选项选项卡上的当前设置变成默认设置。）
- 概要标题标签表示每一页上的第一项的第一、第二、第三和/或第四层的概要标题。
- 页面标题和子标题打印当前的页面标题和子标题。它们是用查看器“插入”菜单上的“新建页面标题”或 TITLE 和 SUBTITLE 命令创建的。如果没有指定任何页面标题或子标题，则该设置将被忽略。

  注意：新页面标题和子标题的字体特征由“选项”对话框（通过选择“编辑”菜单上的“选项”访问）的“查看器”选项卡进行控制。现有页面标题和子标题的字体特征可以通过在查看器中编辑标题进行更改。

要查看页眉和页脚在打印页面上的效果，可选择“文件”菜单上的打印预览。

插入页眉和页脚

▶ 使查看器成为活动窗口（单击该窗口中的任意位置）。

▶ 从菜单中选择：
文件 > 页面属性...

▶ 单击**页眉/页脚**选项卡。

▶ 输入要显示在每一页上的页眉和/或页脚。

页面属性:选项

该对话框控制打印图表大小、打印输出项之间的空格和页编号。

- **打印图表尺寸。**控制相对于定义的页面大小的打印图表的大小。图表的宽高比（宽与高的比）不受打印图表大小的影响。图表的整体打印大小受其高度和宽度的限制。当图表的外边框达到页面的左右边框时，图表大小就无法再扩展以填充额外的页面高度。
- **各项之间的空格。**控制打印的各项之间的空格。每个枢轴表、图表和文本对象都是一个独立的项。该设置不影响各项在查看器中的显示。
- **起始页码。**用指定的数字开始按顺序编排页码。
- **设为默认。**此选项使用此处指定的设置作为新的查看器文档的默认设置。（注意：这使页眉/页脚选项卡和选项选项卡上的当前设置变成默认设置。）

图片 10-13
“页面属性”对话框，“选项”选项卡



更改打印图表大小、页编号和打印项之间的空格

▶ 使查看器成为活动窗口（单击该窗口中的任意位置）。

▶ 从菜单中选择：
文件 > 页面属性...

▶ 单击**选项**选项卡。

▶ 更改设置并单击**确定**。

保存输出

查看器的内容可以保存到查看器文档。保存的文档包括查看器窗口的两个窗格（概要窗格和内容窗格）。

保存查看器文档

▶ 从查看器窗口菜单中选择：
文件 > 保存

▶ 输入文档的名称，然后单击**保存**。

或者，您可以锁定文件以防止在 IBM® SPSS® Smartreader（使用查看器文档的独立产品）中编辑。如果锁定了查看器文档，您可以操作枢轴表（对调行与列，更改显示层等），但无法在 SPSS Smartreader 中编辑任何输出或保存任何更改到查看器文档。该设置对在 IBM® SPSS® Statistics 中打开的查看器文档没有影响。

要用外部格式（例如，HTML 或文本格式）保存结果，请使用“文件”菜单中的“导出”。

章

11

枢轴表

很多结果都以可交互旋转的表形式显示。也就是说，您可以重新排列行、列和层。

注意：此处介绍的多数功能，除创建表时应用的表格外观外，均不适用于轻量表。轻量表无法进行旋转或编辑。

操作枢轴表

操作枢轴表的选项包括：

- 交换行和列
- 移动行和列
- 创建多维层
- 对行和列进行分组和取消分组
- 显示和隐藏行、列和其他信息
- 旋转行标签和列标签
- 查找项的定义

激活枢轴表

在操作或修改枢轴表之前，您需要**激活**表。要激活表：

▶ 双击该表。

或

▶ 右键单击表，从上下文菜单中选择**编辑内容**。

▶ 从子菜单中选择**在查看器中**或**在单独窗口中**。

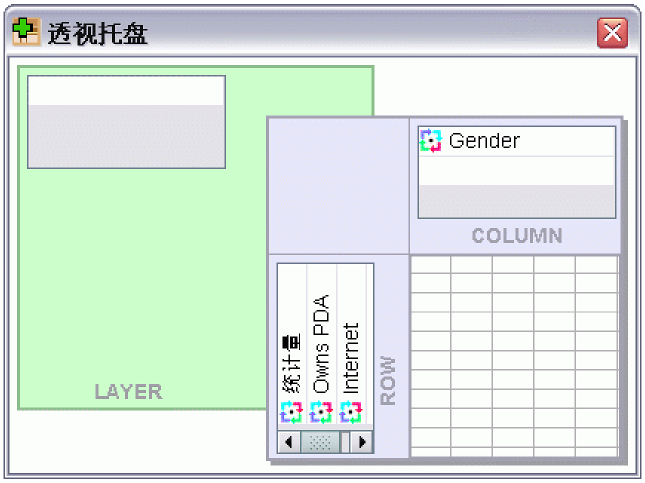
- 默认情况下，双击以激活表的操作将在查看器窗口中激活所有除特大表之外的表。
- 如果需要同时激活多个枢轴表，则必须在单独窗口中激活这些表。

透视表

▶ 激活枢轴表。

▶ 从菜单中选择：
透视 > 透视托盘

图片 11-1
旋转托盘



表包含三个维度：行、列和层。维度中可包含多个元素（或不包含元素）。可以在维度之间或内部移动元素，以更改表的组织。要移动元素，只需将其拖放到所需位置。

更改元素在维度内的显示顺序

要更改表维度（行、列或层）中元素的显示顺序：

▶ 如果旋转托盘尚未打开，则从“枢轴表”菜单中选择：
透视 > 透视托盘

▶ 将维度内的元素拖放到旋转托盘中。

在维度元素中移动行和列

▶ 在表（而非旋转托盘）上，单击您想要移动的行或列的标签。

▶ 将标签拖到新位置。

▶ 从上下文菜单中，选择**插入**或**交换**。

注意：确保“编辑”菜单上的“拖放复制”未启用（选中）。如果启用了“拖放复制”，则取消选择它。

交换行和列

如果只需要转置行和列，可以通过枢轴表采用一种简单方法：

▶ 从菜单中选择：
透视 > 转置行和列

这与将所有行元素拖到列维度，并将所有列元素拖到行维度中的效果相同。

对行或列分组

▶ 选择您想要分组在一起的行或列的标签（单击并拖动，或者按住 Shift 单击以选择多个标签）。

▶ 从菜单中选择：
编辑 > 分组

将自动插入一个组标签。双击组标签可编辑标签文本。

图片 11-2
行和列组及标签

		列组标签		
		女	男	总计
行组标签	职员	206	157	363
	保管员	0	27	27
	经理	10	74	84

注意：要将行或列添加到现有的组，必须先对当前组中的项取消分组。然后创建包含附加项的新组。

对行或列取消分组

▶ 单击您想要取消分组的行或列的组标签的任意位置。

▶ 从菜单中选择：
编辑 > 取消分组

取消分组会自动删除组标签。

旋转行标签或列标签

只能旋转表的最里面列标签和最外面行标签，使其在水平和垂直显示之间切换。

▶ 从菜单中选择：
格式 > 旋转内部列标签

或
格式 > 旋转外部行标签

图片 11-3
已旋转的列标签

		频率	百分比	有效百分比	累积百分比
有效	未完成高中	460	54.1	54.1	54.1
	高中	235	27.6	27.6	81.8
	大专				
	大学				
	研究生				
	合计				

		频率	百分比	有效百分比	累积百分比
有效	未完成高中	460	54.1	54.1	54.1
	高中	235	27.6	27.6	81.8
	大专	101	11.9	11.9	93.6
	大学	49	5.8	5.8	99.4
	研究生	5	.6	.6	100.0
	合计	850	100.0	100.0	

只能旋转最里面的列标签和最外面的行标签。

使用层

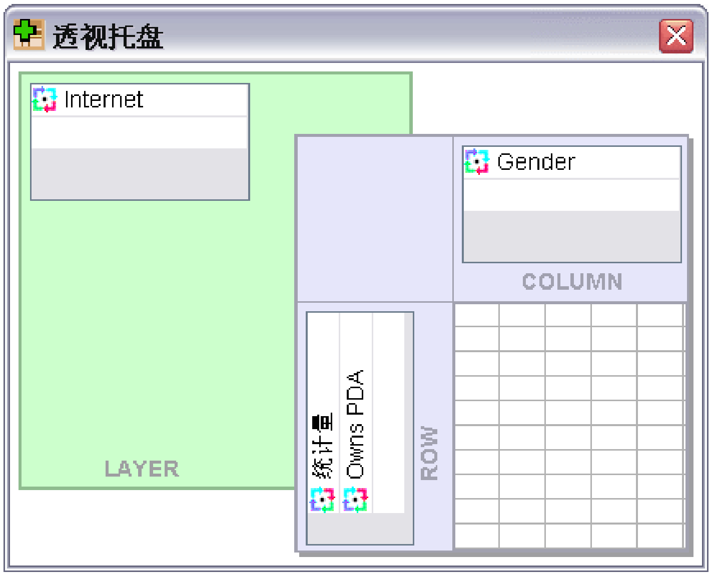
您可以为每个类别或类别的组合显示一个单独的二维表。该表可视为按层堆积，只有最顶层可视。

创建并显示层

要创建层：

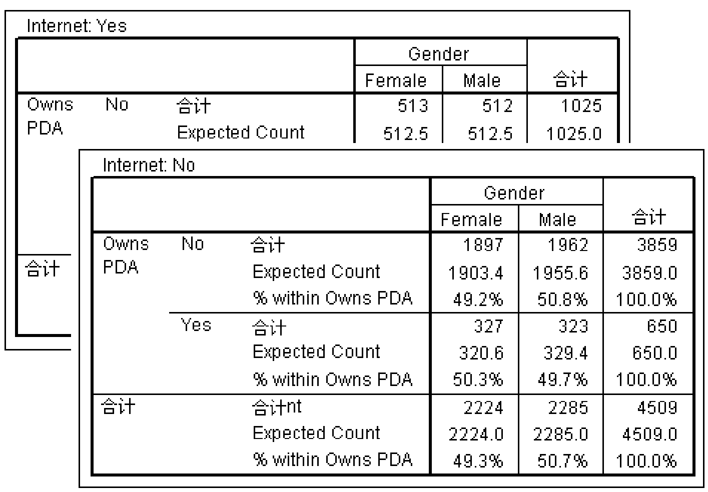
- 激活枢轴表。
- 如果旋转托盘尚未打开，则从“枢轴表”菜单中选择：
  透视 > 透视托盘
- 将一个元素从行维度或列维度拖到层维度。

图片 11-4
将类别移动到层中



将元素移至层维度中，会创建一个多维表，但只会显示单个二维分区。可视的表是顶层的表。例如，如果在层维度中有一个是/否类别变量，则多维表将有两层。它们分别对应于是类别和否类别。

图片 11-5
单独层中的类别



Internet: Yes

			Gender		合计
			Female	Male	
Owns PDA	No	合计	513	512	1025
		Expected Count	512.5	512.5	1025.0
合计					

Internet: No

			Gender		合计
			Female	Male	
Owns PDA	No	合计	1897	1962	3859
		Expected Count	1903.4	1955.6	3859.0
		% within Owns PDA	49.2%	50.8%	100.0%
	Yes	合计	327	323	650
		Expected Count	320.6	329.4	650.0
		% within Owns PDA	50.3%	49.7%	100.0%
合计		合计nt	2224	2285	4509
		Expected Count	2224.0	2285.0	4509.0
		% within Owns PDA	49.3%	50.7%	100.0%

更改显示层

▶ 从层的下拉列表中选择一个类别（在枢轴表本身而非旋转托盘上）。

图片 11-6
从下拉列表中选择层

Internet: No (No / Yes)

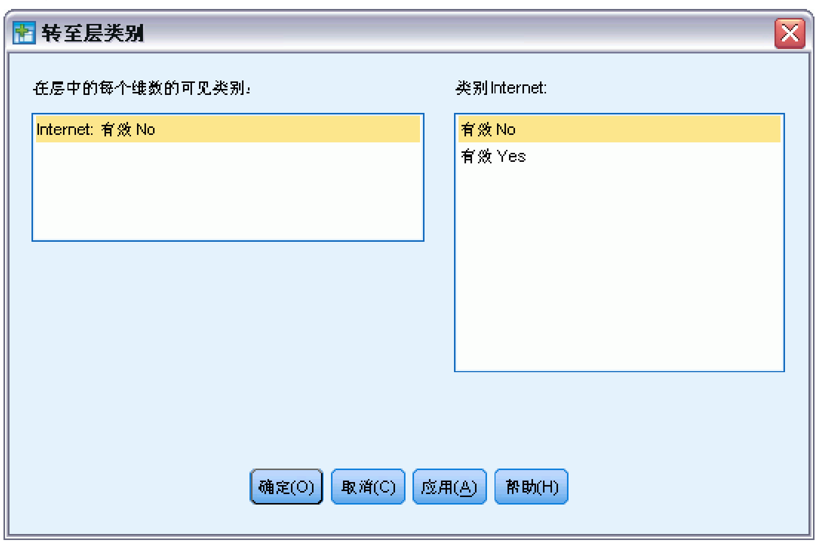
			Gender		合计
			Female	Male	
Owns PDA	No	合计	1897	1962	3859
		Expected Count	1903.4	1955.6	3859.0
		% within Owns PDA	49.2%	50.8%	100.0%
	Yes	合计	327	323	650
		Expected Count	320.6	329.4	650.0
		% within Owns PDA	50.3%	49.7%	100.0%
合计		合计	2224	2285	4509
		Expected Count	2224.0	2285.0	4509.0
		% within Owns PDA	49.3%	50.7%	100.0%

转至层类别

“转到层类别”使您可以在枢轴表中更改层。当存在许多层或选定层包含许多类别时，此对话框尤其有用。

▶ 从菜单中选择：
透视 > 转到层...

图片 11-7
“转到层类别”对话框



▶ 在“可视类别”列表中选择一个层维度。“类别”列表将显示选定维度的所有类别。

▶ 在“类别”列表中，选择所需类别，然后单击**确定**或**应用**。

显示和隐藏项目

许多类型的单元格可以隐藏，包括：

■ 维度标签

- 类别，包括行或列中的标签单元格和数据单元格
- 类别标签（不隐藏数据单元格）
- 脚注、标题和题注

隐藏表中的行和列

▶ 右键单击您想要隐藏的行或列的类别标签。

▶ 从上下文菜单中选择：
选择 > 数据和标签单元格

▶ 再次右键单击类别标签，从上下文菜单中选择隐藏类别。

或

▶ 从“视图”菜单中选择隐藏。

显示表中的隐藏行和列

▶ 右键单击与隐藏的行或列在相同维度中的另一行或列。

▶ 从上下文菜单中选择：
选择 > 数据和标签单元格

▶ 从菜单中选择：
视图 > 显示[维名称]中的所有类别

或

▶ 要显示激活的枢轴表中的所有隐藏的行和列，从菜单中选择：
视图 > 全部显示

这将显示表中的所有隐藏行和列。（如果在此表的“表格属性”中选择了隐藏空行列，则完全为空的行或列仍会隐藏。）

隐藏和显示维度标签

▶ 选择维度标签或者维度中的任何类别标签。

▶ 从“查看”菜单或上下文菜单中，选择隐藏维度标签或显示维度标签。

隐藏和显示表标题

要隐藏标题：

▶ 选择标题。

▶ 从“视图”菜单中选择隐藏。

要显示隐藏的标题：

▶ 从“视图”菜单中选择**显示全部**。

表格外观

表格外观是定义表的外观的一组属性。您可以选择先前定义的表格外观，或创建自己的表格外观。

- 在应用表格外观前后，可以使用单元格属性更改个别单元格或单元格组的单元格格式。即使您应用新的表格外观，编辑的单元格格式仍将保持不变。
- 您还可以选择将所有单元格重置为由当前表格外观定义的单元格格式。这会重置已编辑的所有单元格。如果在“表格外观文件”列表中选择了**按显示**，则任何已编辑的单元格都将重置为当前的表格属性。

注意：以 IBM® SPSS® Statistics 的早期版本创建的表格外观不能用于 16.0 或更高版本。

应用或保存表格外观

▶ 激活枢轴表。

▶ 从菜单中选择：
格式 > 表格外观...

图片 11-8
“表格外观”对话框



▶ 从文件列表中选择一个表格外观。要从另一个目录中选择文件，请单击**浏览**。

▶ 单击**确定**将表格外观应用到选定的枢轴表。

编辑或创建表格外观

▶ 在“表格外观”对话框中，从文件列表中选择一个表格外观。

▶ 单击**编辑外观**。

▶ 为需要的属性调整表格属性，然后单击**确定**。

▶ 单击**保存外观**保存已编辑的表格外观，或者单击**另存为**将其保存为新的表格外观。

编辑表格外观仅影响选定的枢轴表。已编辑的表格外观不会应用到使用该表格外观的任何其他表，除非您选择这些表并重新应用表格外观。

表格属性

“表格属性”允许您设置表的常规属性，设置表的各个部分的单元格样式，以及将一组属性保存为表格外观。您可以：

- 控制常规属性，例如隐藏空行列，以及调整打印属性。
- 控制脚注标记的格式和位置。
- 确定数据区域中的单元格的特定格式、行和列标签的特定格式以及表的其他区域的特定格式。
- 控制形成表的每个区域的边框的线条的宽度和颜色。

更改枢轴表属性

▶ 激活枢轴表。

▶ 从菜单中选择：
格式 > 表格属性...

▶ 选择一个选项卡（**常规**、**脚注**、**单元格格式**、**边框**或**打印**）。

▶ 选择您想要的选项。

▶ 单击**确定**或**应用**。

新的属性将应用到选定的枢轴表。要将新的表格属性应用到表格外观而不仅仅是选定的表，请编辑表格外观（“格式”菜单中的“表格外观”）。

表格属性：一般

多个属性作为一个整体应用到表。您可以：

- 显示或隐藏空的行与列。（空的行或列在任何数据单元格中均不存在任何内容。）
- 控制在长表中显示的默认行数。要在表中显示所有行，而不考虑表长度，请取消选中**按行显示表格**。
- 控制行标签的位置，它们可以位于左上角或嵌套。
- 控制最大和最小列宽（以磅表示）。

图片 11-9
“表格属性”对话框，“常规”选项卡



要更改常规表格属性：

▶ 单击常规选项卡。

▶ 选择您想要的选项。

▶ 单击确定或应用。

设置待显示行

默认情况下，行数较多的表显示在多个部分（每个部分 100 行）中。控制在表中显示的行数：

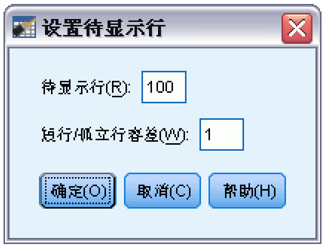
▶ 选择按行显示表格。

▶ 单击设置待显示行。

或

▶ 从已激活枢轴表的“视图”菜单中，选择按行显示表格和设置待显示行。

图片 11-10
“设置待显示行”对话框



待显示行。 控制一次显示的最大行数。导航控件允许您移动到表格的不同部分。最小值为 10。默认值为 100。

短行/孤立行容差。 控制表格最内层行维度的最大行数，以便在表格的显示视图之间拆分。例如，如果最内层行维度的每个组中有 6 个类别，则指定值 6 可以防止任何组在显示视图之间拆分。该设置可能导致某个显示视图中的总行数超过指定的最大待显示行数。

图片 11-11
带有默认容差的显示行

行 1 - 10 / 24

				计数	Percent
Income category in thousands	Under $25	Job satisfaction	Highly dissatisfied	350	29.8%
			Somewhat dissatisfied	316	26.9%
			Neutral	250	21.3%
			Somewhat satisfied	157	13.4%
			Highly satisfied	101	8.6%
			总计	1174	100.0%
	$25 - $49	Job satisfaction	Highly dissatisfied	531	22.2%
			Somewhat dissatisfied	534	22.4%
			Neutral	542	22.7%
			Somewhat satisfied	496	20.8%

图片 11-12
容差设置基于内层行维度组中的行数

行 1 - 12 / 24

				计数	Percent
Income category in thousands	Under $25	Job satisfaction	Highly dissatisfied	350	29.8%
			Somewhat dissatisfied	316	26.9%
			Neutral	250	21.3%
			Somewhat satisfied	157	13.4%
			Highly satisfied	101	8.6%
			总计	1174	100.0%
	$25 - $49	Job satisfaction	Highly dissatisfied	531	22.2%
			Somewhat dissatisfied	534	22.4%
			Neutral	542	22.7%
			Somewhat satisfied	496	20.8%
			Highly satisfied	285	11.9%
			总计	2388	100.0%

表格属性：脚注

脚注标记的属性包括样式和相对于文本的位置。

- 脚注标记的样式为数字（1、2、3...）或字母（a、b、c...）。
- 脚注标记可以作为上标或下标附加在文本上。

图片 11-13
“表格属性”对话框，“脚注”选项卡



要更改脚注标记属性：

▶ 单击**脚注**选项卡。

▶ 选择脚注数字格式。

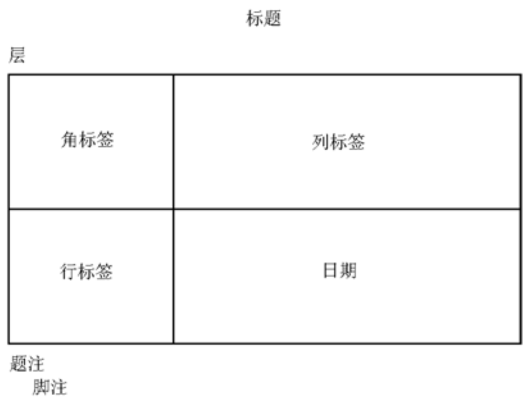
▶ 选择标记位置。

▶ 单击**确定**或**应用**。

表格属性：单元格格式

表可以按不同的区域设置格式：标题、层、角标签、行标签、列标签、数据、题注和脚注。对于表的每个区域，您都可以修改相关联的单元格格式。单元格格式包括文本特征（如字体、大小、颜色和样式）、水平和垂直对齐方式、背景色以及内单元格边距。

图片 11-14
表的区域



单元格格式应用到区域（信息的类别）。它们不是个别单元格的特征。当对表进行旋转时，此区别是重要的考虑因素。

例如，

- 如果指定粗体作为列标签的单元格格式，则不管列维度中当前显示的信息是什么，列标签总是显示为粗体。如果您将一项从列维度移至其他维度，它不会保留列标签的粗体特征。
- 如果您仅是通过突出显示已激活的枢轴表中的单元格，并单击工具栏上的“粗体”按钮而将列标签变为粗体，则不管您将那些单元格移至哪个维度，其内容总是保留粗体，而对于其他移至列维度的项，列标签不会保留粗体特征。

图片 11-15
“表格属性”对话框，“单元格格式”选项卡



要更改单元格格式：

- 选择**单元格格式**选项卡。
- 从下拉列表中选择一个区域，或单击样本的一个区域。
- 选择该区域的特征。您的选择会反映在样本中。
- 单击**确定**或**应用**。

交替行颜色

应用不同的背景和/或文本颜色到表格数据区域中的交替行：

- 从“区域”下拉列表中选择**数据**。
- 选择（选中）背景颜色组中的**交替行颜色**。
- 选择用于交替行背景和文本的颜色。

交替行颜色仅影响表格的数据区域。它们不会影响行或列标签区域。

表格属性：边框

对于表中的每个边框位置，您都可以选择线样式和颜色。如果选择**无**作为线样式，则在选定位置不会出现线条。

图片 11-16
“表格属性”对话框，“边框”选项卡



要更改表边框：

▶ 单击**边框**选项卡。

▶ 选择边框位置，方法是在列表中单击边框的名称，或者在“样本”区域中单击线。

▶ 选择一种线样式或选择**无**。

▶ 选择一种颜色。

▶ 单击**确定**或**应用**。

表格属性：打印

您可以控制所打印枢轴表的以下属性：

- 打印表的所有层或仅打印顶层，并在单独页面上打印每一层。
- 水平或垂直地收缩表，以适合页面打印。
- 如果表对于定义的页面大小太宽和/或太长，则通过控制包含在表的任何打印部分的最小行数和列数来控制短行/孤立行。

注意：如果当前页面由于在表的上方包含其他输出而导致无法容纳整个表，但表能容纳于定义的页面长度，则会自动将该表打印在新的一页中，而不管短行/孤立行设置如何。

■ 包含无法容纳在单页中的表的接续文本。您可以在每页的底部和顶部显示接续文本。如果未选择任一选项，则不会显示接续文本。

图片 11-17
“表格属性”对话框，“打印”选项卡



要控制枢轴表打印属性：

▶ 单击**打印**选项卡。

▶ 选择您想要的打印选项。

▶ 单击**确定**或**应用**。

单元格属性

单元格属性会应用到选定的单元格。您可以更改字体、值的格式、对齐方式、边距和颜色。单元格属性会覆盖表格属性；所以，更改表格属性不会影响到任何单独应用的单元格属性。

要更改单元格属性：

▶ 激活一个表并在表中选择单元格。

▶ 从“格式”菜单或上下文菜单中选择**单元格属性**。

字体及背景

“字体及背景”选项卡控制表中选定单元格的字体样式、颜色和背景颜色。

图片 11-18
“单元格属性”对话框，“字体及背景”选项卡



格式值

“格式值”选项卡控制选定单元格的值格式。您可以选择数字、日期、时间或货币格式，还可以调整所显示的小数位数。

图片 11-19
“单元格属性”对话框，“格式值”选项卡



注意：货币格式列表包含美元格式（带有前导美元符号的数字）和五个自定义货币格式。默认情况下，所有自定义货币格式设置为默认的数值格式，其中不包含货币或其他自定义符号。有关定义自定义货币格式的信息，请参见货币选项。

对齐与边缘

“对齐与边缘”选项卡控制选定单元格的值的水平和垂直对齐方式，以及上下左右边距。混合水平对齐按照单元格的类型对齐其内容。例如，日期为右对齐，而文本值为左对齐。

图片 11-20
“单元格属性”对话框，“对齐与边缘”选项卡



脚注和题注

您可以向表添加脚注和题注。还可以隐藏脚注或题注，更改脚注标记，以及对脚注重新编号。

某些脚注属性由表特性控制。

添加脚注和题注

要向表添加题注：

▶ 从“插入”菜单中选择**题注**。

可以向表中的任何项添加脚注。要添加脚注：

▶ 单击已激活的枢轴表的标题、单元格或题注。

▶ 从“插入”菜单中选择**脚注**。

隐藏或显示题注

要隐藏题注：

- 选择题注。
- 从“视图”菜单中选择**隐藏**。

要显示隐藏的题注：

- 从“视图”菜单中选择**显示全部**。

隐藏或显示表中的脚注

要隐藏脚注：

- 选择脚注。
- 从“查看”菜单中选择**隐藏**，或从上下文菜单中选择**隐藏脚注**。

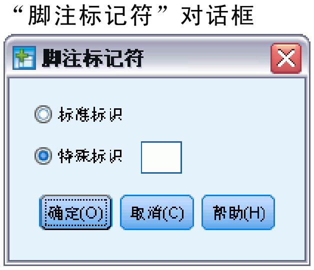
要显示隐藏的脚注：

- 从“视图”菜单中选择**显示全部脚注**。

脚注标记符

“脚注标记”更改用于标记脚注的字符。

图片 11-21
“脚注标记符”对话框



要更改脚注标记：

- 选择脚注。
- 从“格式”菜单中选择**脚注标记**。
- 输入一个或两个字符。

对脚注重新编号

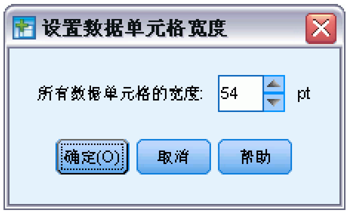
当通过交换行、列和层对表进行旋转之后，可能会打乱脚注的顺序。对脚注重新编号：

- 从“格式”菜单中选择**对脚注重新编号**。

数据单元格宽度

“设置数据单元格宽度”用于将所有数据单元格设置为相同的宽度。

图片 11-22
“设置数据单元格宽度”对话框



要为所有数据单元格设置宽度：

▶ 从菜单中选择：
格式 > 设置数据单元格宽度...

▶ 输入一个值作为单元格宽度。

更改列宽

▶ 单击并拖动列边框。

显示枢轴表中的隐藏边框

对于不含很多可视边框的表，可以显示隐藏边框。这可以简化更改列宽之类的任务。

▶ 从“视图”菜单中选择**网格线**。

图片 11-23
显示隐藏边框的网格线

Internet				Gender: Female	Gender: Male	Total
No	Owns PDA	No	Count	1897%	1962%	3859%
			Expected Count	1903.4%	1955.6%	3859.0%
			% within Owns PDA	49.2%	50.8%	100.0%
		Yes	Count	327%	323%	650%
	Total					
Yes	Owns Pl					
	Total					

Internet				Gender: Female	Gender: Male	Total
No	Owns PDA	No	Count	1897%	1962%	3859%
			Expected Count	1903.4%	1955.6%	3859.0%
			% within Owns PDA	49.2%	50.8%	100.0%
		Yes	Count	327%	323%	650%
			Expected Count	320.6%	329.4%	650.0%
			% within Owns PDA	50.3%	49.7%	100.0%
	Total		Count	2224%	2285%	4509%
			Expected Count	2224.0%	2285.0%	4509.0%
			% within Owns PDA	49.3%	50.7%	100.0%
Yes	Owns PDA	No	Count	513%	512%	1025%
			Expected Count	512.5%	512.5%	1025.0%
			% within Owns PDA	50.0%	50.0%	100.0%
		Yes	Count	305%	306%	611%
			Expected Count	305.5%	305.5%	611.0%
			% within Owns PDA	49.9%	50.1%	100.0%
	Total		Count	818%	818%	1636%
			Expected Count	818.0%	818.0%	1636.0%
			% within Owns PDA	50.0%	50.0%	100.0%

在枢轴表中选择行和列

在枢轴表中，对如何选择整行和整列以及指示选定的行或列可以跨不连续的表区域的可视突出显示部分有一些限制。要选择整层或整列：

▶ 单击行或列标签。

▶ 从菜单中选择：
编辑 > 选择 > 数据和标签单元格

或

▶ 右键单击行或列的类别标签。

▶ 从上下文菜单中选择：
选择 > 数据和标签单元格

或

▶ 按住 Ctrl 和 Alt，并单击行或列标签。

打印枢轴表

有多个因素会影响打印枢轴表外观的方式，可以通过更改枢轴表属性控制这些因素。

- 对于多维枢轴表（带有多层的表），您可以打印所有层，也可以只打印顶层（可视层）。
- 对于长或宽的枢轴表，可以自动调整表的大小以适合页面，也可以控制表分隔符和分页符的位置。
- 如果有的表太宽或太长而无法在单个页面中打印，您可以控制页间的表分隔符位置。

可以使用“文件”菜单上的“打印预览”查看打印的枢轴表的外观。

控制宽表和长表的表分隔符

有的枢轴表太宽或太长，无法在定义的页面大小中打印，程序会自动分割它们，并将它们打印在多个部分中。（对于宽表，如果空间足够，多个部分会打印在同一页上。）您可以：

- 控制分割大型表的行或列的位置。
- 指定分割表时应保留在一起的行和列。
- 调整大型表的大小以适合定义的页面大小。

指定打印的枢轴表的行和列分隔符

▶ 单击想要插入分隔符的位置左边的列标签，或者单击想要插入分隔符的位置上面的行标签。

▶ 从菜单中选择：
格式 > 在此处分隔

指定保留在一起的行或列

▶ 选择要保留在一起的行或列的标签。（单击并拖动，或者按住 Shift 单击以选择多个行或列标签。）

▶ 从菜单中选择：
格式 > 保留在一起

从枢轴表创建图表

▶ 双击激活枢轴表。

▶ 选择想要在图中显示的行、列或单元格。

▶ 右键单击选定区域中的任意位置。

▶ 从上下文菜单中选择**创建图形**并选择一个图表类型。

轻量表

表可以呈现为全功能枢轴表或轻量表。轻量表可以比全功能枢轴表更快呈现，但它们不支持许多透视和编辑功能，包括：

- 无法透视轻量表。您可以显示多层表的不同层，但无法在维度之间移动表元素（例如，交换行和列）或更改行或列的显示位置。
- 无法编辑轻量表。您无法隐藏行或列、更改列宽度、修改数据值、更改单元格或表格属性，也无法应用其他表格外观。
- 轻量表不支持在表格外观中定义的自定义维数边界属性。所有其他表格外观属性会在创建表格时被接受。
- 由于无法编辑轻量表，“查找和替换”对话框中的“替换”功能将不会替换轻量表中的指定值。

要将轻量表转换为全功能枢轴表：

▶ 右键单击“查看器”窗口中的表格或表格编辑窗口。

▶ 从上下文菜单中，选择**启用编辑**

有关如何创建轻量表以取代全功能枢轴表的详细信息，请参阅 枢轴表选项

章

12

模型

一些结果显示为模型，在输出查看器中显示为一种特殊类型的直观表示。在输出查看器中显示的直观表示不是唯一可用的模型视图。单个模型包含许多不同视图。您可以激活模型查看器中的模型，并直接与模型进行交互作用，以显示可用模型视图。您还可选择打印并导出模型中的所有视图。

与模型进行交互作用

与模型进行交互作用，请首先将其激活：

▶ 双击该模型。

或

▶ 右键单击模型，从上下文菜单中选择**编辑内容**。

▶ 从子菜单中选择**在单独窗口中**。

激活模型将在模型查看器中显示模型。

使用模型查看器

模型查看器是一种显示可用模型视图和编辑模型视图外观的交互式工具。（有关显示模型查看器的信息，请参见与模型进行交互作用第 236 页码。）有两种不同样式的模型查看器：

- **分割成主/辅助视图。** 在此样式中，**主视图**出现在模型查看器的左部。主视图显示模型的一些一般直观表示（例如网络图片）。主视图本身可有多个模型视图。主视图下方的下拉列表允许您从可用主视图中选择。

  辅助视图出现在模型查看器的右部。与主视图中的一般直观表示相比，辅助视图通常显示模型的更多详细直观表示（包括表格）。与主视图一样，辅助视图可有多个模型视图。辅助视图下方的下拉列表允许您从可用主视图中选择。辅助视图还可为主视图中选择的元素显示特定直观表示。例如，根据模型类型，您可以在主视图中选择一个变量节点，以在辅助视图中显示该变量的表格。
- **通过缩略图，一次打开一个视图。** 在此样式中，只有一个视图可见，而其他视图则通过模型查看器左部的缩略图进行访问。每个视图显示模型的一些直观表示。

显示的特定直观表示取决于创建模型的过程。有关使用特定模型的信息，请参考创建模型的过程文档。

模型视图表格

在模型查看器中显示的表格不是枢轴表。您无法像操作枢轴表一样操作这些表格。

设置模型属性

在模型查看器中，您可以设置特定模型属性。

复制模型视图

您还可复制模型查看器中的单个模型视图。

模型属性

从模型查看器中的菜单选择：

文件 > 属性

每个模型都有相关属性，让您指定从输出查看器中打印哪些视图。缺省情况下，只打印在输出查看器中可见的视图。一般总是主视图，且只有一个主视图。您还可指定打印所有可用模型视图。这些包括所有主视图以及所有辅助试图（基于主视图选择的辅助视图除外；这些是不可打印的）。注意，您还可打印模型查看器自带的单个模型视图。

复制模型视图

从模型查看器的“编辑”菜单中，您可以复制当前显示的主视图或当前显示的辅助视图。只复制一个模型视图。您可以将模型视图粘贴到输出查看器中，在这里单个模型视图随后以可以在图形板编辑器中编辑的直观表示显示。“粘贴到输出查看器”允许您同时显示多个模型视图。您还可粘贴到其他应用程序中，视图会根据目标应用程序显示为一个图像或表格。

打印模型

从模型查看器中打印

您可以打印模型查看器自带的单个模型视图。

- ▶ 在模型查看器中激活模型。
- ▶ 从菜单中选择：

  视图 > 编辑模式
- ▶ 在主视图或辅助视图（根据您想打印哪一个）中的“一般工具栏”调色板上，单击打印图标。（如果此调色板未显示，请从“视图”菜单中选择**调色板>一般**。）

从输出查看器中打印

当您从输出查看器中打印时，为特定模型所打印的视图数取决于模型属性。模型可以设置为只打印显示的视图或所有可用模型视图。

探索模型

默认情况下，当您从输出查看器中导出模型时，包括或排除模型视图由每个模型的模型属性控制。有关模型属性的更多信息，请参见模型属性第 237 页码。在导出时，您可以覆盖此设置，并包括所有模型视图或只包括当前可见的模型视图。在“导出输出”对话框中，单击“文档”组中的**更改选项**...。有关导出和此对话框的更多信息，请参见导出输出第 197 页码。请注意，包括表在内的所有模型视图，均导出为图形。还请注意，基于主视图选择的辅助视图不可导出。

将模型中使用的字段保存到新的数据集

您可以将模型中使用的字段保存到新的数据集。

▶ 在模型查看器中激活模型。

▶ 从菜单中选择：
生成 > 字段选择（模型输入和目标）

图片 12-1
基于模型中字段的字段选择



数据集名称。 请指定一个有效的数据集名称。可以在同一会话中继续使用数据集，但不会将其另存为文件，除非在会话结束之前明确将其保存为文件。数据集名称必须符合变量命名规则。

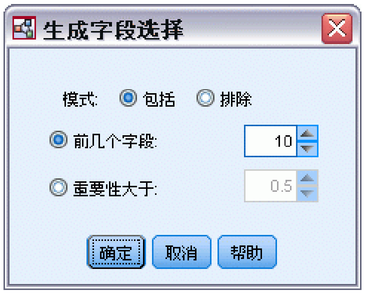
根据重要性将预测变量保存到新的数据集

您可以根据预测变量重要性图表中的信息，将预测变量保存到新的数据集。

▶ 在模型查看器中激活模型。

▶ 从菜单中选择：
生成 > 字段选择（预测变量重要性）

图片 12-2
基于重要性的字段选择



变量的最大数。包括或排除等于指定数量的最重要预测变量。

重要性大于。包括或排除所有相对重要性高于指定值的预测变量。

▶ 单击确定后，将显示“新数据集”对话框。

图片 12-3
字段选择：新数据集

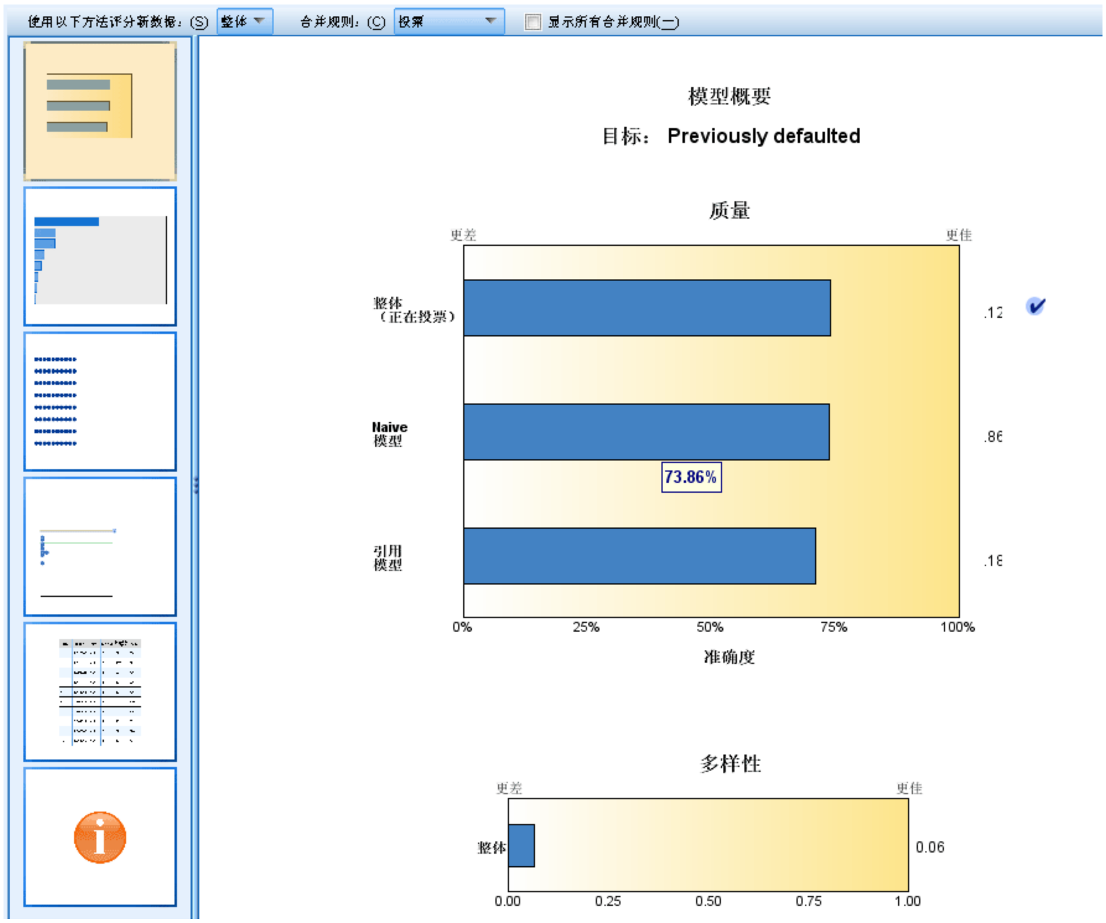


数据集名称。 请指定一个有效的数据集名称。可以在同一会话中继续使用数据集，但不会将其另存为文件，除非在会话结束之前明确将其保存为文件。数据集名称必须符合变量命名规则。

整体模型

整体模型提供了有关整体中的组件模型和整体性能的信息。

图片 12-4
模型摘要视图



主（独立视图）工具栏允许您选择使用整体或参考模型来进行评分。如果使用整体进行评分，您还可以选择组合规则。这些更改不需要重新执行模型；但是，这些选择将保存到模型 以供评分和/或下游模型评估。它们也会影响从整体查看器导出的 PMML。

组合规则。在对整体评分时，此规则用于组合来自基本模型的预测值，以计算整体得分值。

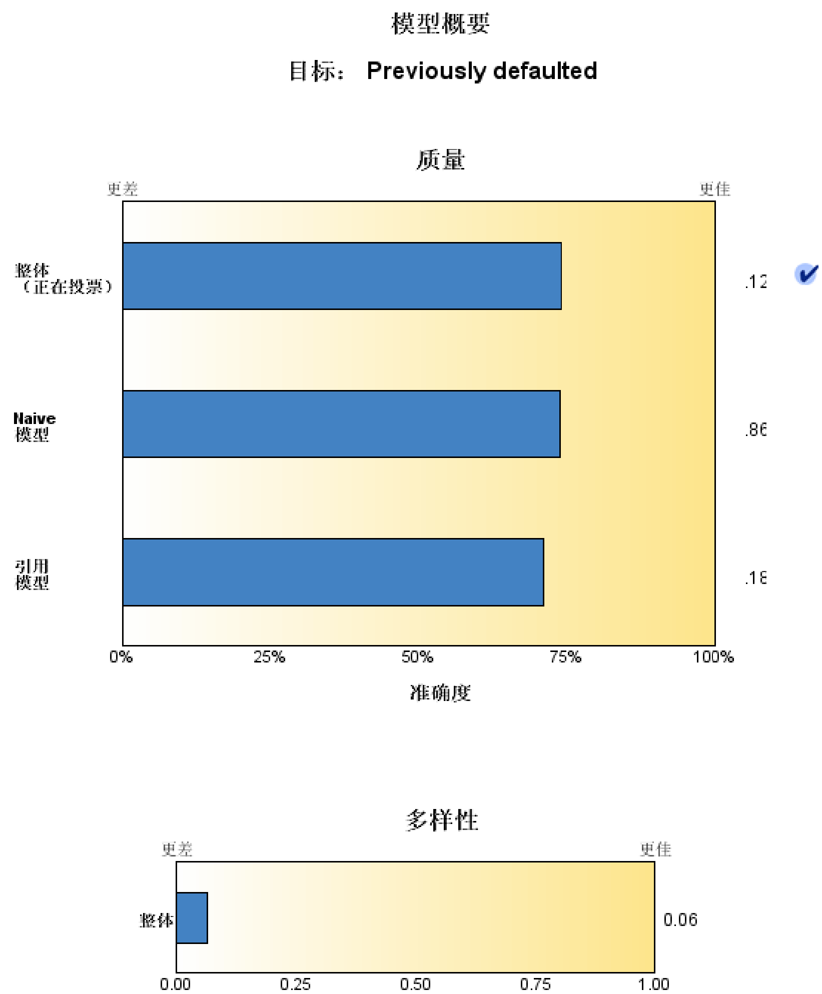
- 可以使用投票、最高概率或最高平均概率来组合**分类**目标的整体预测值。**投票**选择在基本模型中最常具有最高概率的类别。**最高概率**选择在所有基本模型中达到单一最高概率的类别。**最高平均概率**选择当类别概率在基本模型中取平均值时具有最高值的类别。
- 可以通过对来自基本模型的预测值取平均值或中位数，对**连续**目标的整体预测值进行组合。

默认值取自在建模过程中生成的指定。更改组合规则会重新计算模型精确性并更新模型精确性的所有视图。也会更新预测变量重要性图表。如果选择参考模型用于评分，则此控件将被禁用。

显示所有组合规则。 选择该选项时，所有可用组合规则的结果将显示在模型质量图表中。组件模型精确性图表也将更新以显示每种投票方式的参考线。

模型摘要

图片 12-5
模型摘要视图



“模型摘要”视图是整体质量和差异性的快照摘要。

质量。该图表显示与参考模型和 navie 模型相比较的最终模型精确性。精确性越大，模型越好的格式；“最佳”模型将具有最高精确性。对于分类目标，精确性就是预测值与观测值匹配的记录百分比。对于连续目标，精确性为 1 减去预测中的平均绝对误差（预测值的绝对平均值减去观测值）与预测值范围（最大预测值减去最小预测值）的比率。

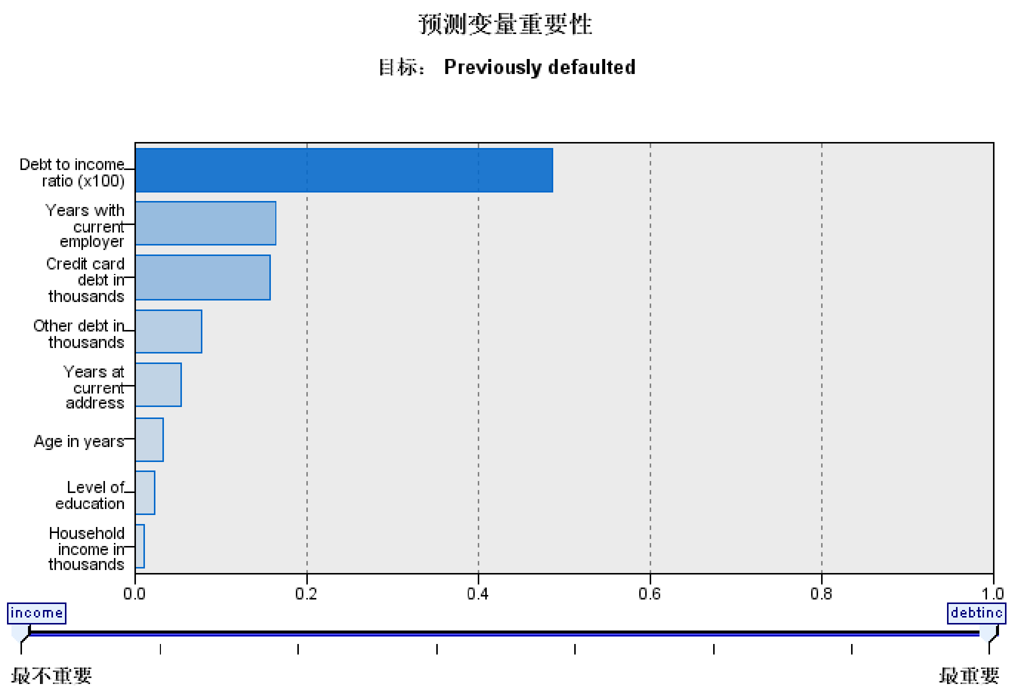
对于 bagging 整体，参考模型是构建在整个培训分区上的标准模型。对于 boosted 整体，参考模型是第一个组件模型。

如果未构建模型，则由 Naive 模型代表精确性，并将所有记录分配给模态类别。不会为连续目标计算 Naive 模型。

差异性。该图表显示用于构建整体的组件模型间的“观点差异性”，以越大则差异性越大格式表示。这是一种基本模型间预测差异程度的测量。差异性对 boosted 整体模型不可用，同时也不会对连续目标显示。

预测变量重要性

图片 12-6
预测变量重要性视图

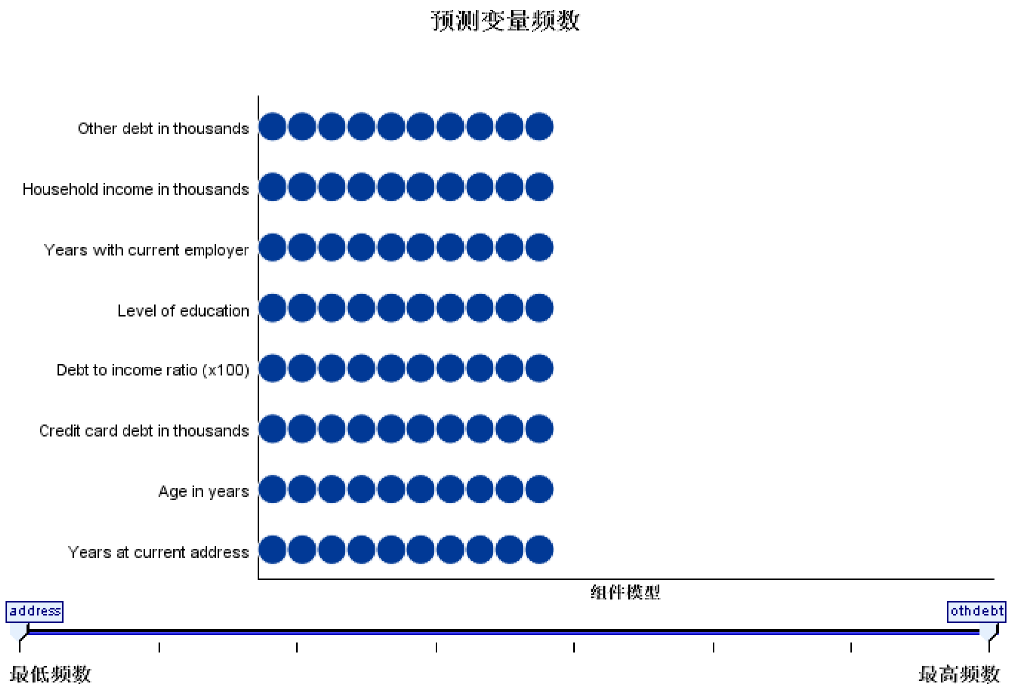


通常，您需要将建模工作专注于最重要的预测变量字段，并考虑删除或忽略那些最不重要的预测变量字段。预测变量重要性图表可以在模型估计中指示每个预测变量的相对重要性，从而帮助您实现这一点。由于它们是相对值，因此显示的所有预测变量的值总和为 1.0。预测变量重要性与模型精度无关。它只与每个预测变量在预测中的重要性有关，而不涉及预测是否精确。

预测变量重要性对所有整体模型均不可用。预测变量集在组件模型之间可能会有所不同，但可以为至少在一个组件模型中使用的预测变量计算重要性。

预测变量频率

图片 12-7
预测变量频率视图

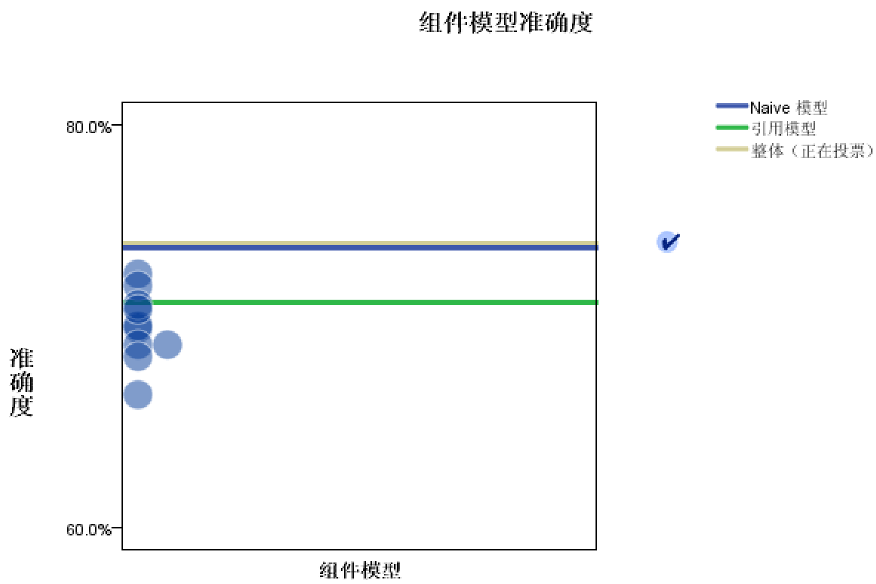


由于选择的建模方法或预测变量选择不同，预测变量集在组件模型间也可能不同。预测变量频率图是一个点图，显示了预测变量在整体组件模型中的分布。每个点代表包含预测变量的一个或多个组件模型。预测变量绘制在 y 轴上，并以频率的降序排序；因此，最顶端的是在最多组件模型中使用的预测变量，而最低端的是在最少组件模型中使用的预测变量。将显示排在前 10 位的预测变量。

出现频率最高的预测变量通常是最重要的。此图对于使预测变量集在组件模型间保持一致的方法没用。

组件模型精确性

图片 12-8
组件模型精确性视图



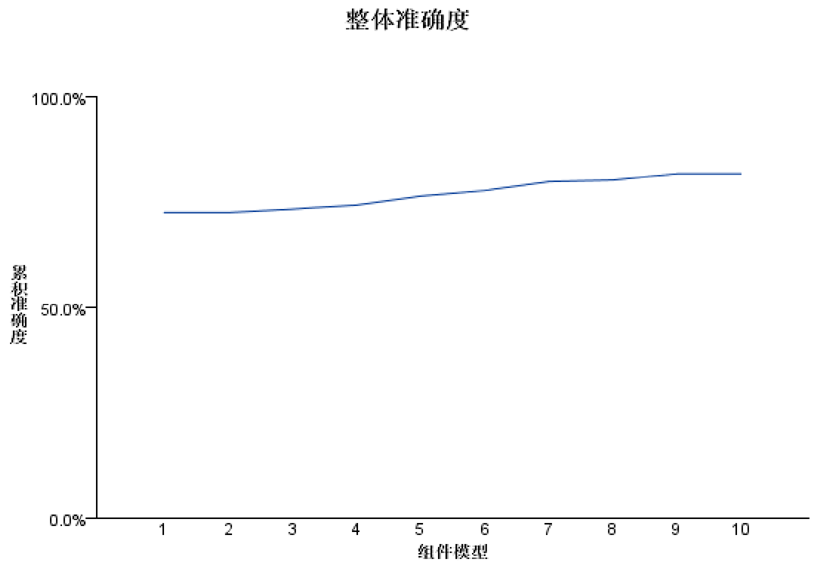
该图表是组件模型预测精确性的点图。每个点代表在 y 轴上绘制了精确性水平的一个或多个组件模型。悬停在任意点上可获得对应的单独组件模型的信息。

参考线。该图显示整体的颜色编码线以及参考模型和 naïve 模型。对应于要用于评分的模型的线的旁边会显示一个复选标记。

互动。该图表会在您更改组合规则时更新。

Boosted 整体。为 boosted 整体显示一个线图。

图片 12-9
整体精确性视图，boosted 整体



组件模型详细信息

图片 12-10
组件模型详细信息视图

组件模型详细信息

模型	准确度	方法	预测变量	模型大小（突触）	记录
1	72.6%		8	77	700
2	72.0%		8	107	692
3	69.9%		8	92	708
4	70.0%		8	107	685
5	71.1%		8	107	706
6	69.1%		8	92	690
7	69.1%		8	92	696
8	70.8%		8	122	703
9	66.6%		8	62	726
10	68.5%		8	107	701

该表显示关于组件模型的信息，按行列出。默认情况下，组件模型按模型编号的升序排序。您可以按任意列的值对这些行进行升序或降序排序。

模型。代表组件模型创建顺序的数字。

精确性。百分比形式的整体精确性。

方法。建模方法。

预测变量。组件模型中使用的预测变量数。

模型大小。模型大小取决于建模方法：对于树模型，它是树中节点的数量；对于线性模型，它是系数的数量；对于神经网络，它是神经元的数量。

记录。训练样本中输入记录的加权数量。

自动数据准备

图片 12-11
自动数据准备视图

自动数据准备

目标： Total sales

字段	角色	采取的操作
Age category	预测变量	合并类别使与目标的关联最大化
Primary keyword set	预测变量	合并类别使与目标的关联最大化
Promotion	预测变量	将测量等级从连续改为顺序
Secondary keyword set	预测变量	合并类别使与目标的关联最大化

如果原始字段名称是 X，则变换后的字段名称是 X_transformed。从分析中排除原始字段，但包括了变换后的字段。

此视图显示在自动数据准备 (ADP) 步骤中排除了哪些字段，以及转换字段的派生方式等信息。对于每个转换或排除字段，在此表中列出了字段名、在分析中的角色，以及 ADP 步骤所采取的操作。这些字段按其名称的字母升序排列。

操作 Trim outliers（如果显示）表示位于截断值（平均值的 3 个标准差）之外的连续预测变量值被设为截断值。

拆分模型查看器

拆分模型查看器列出了每个拆分的模型，并提供了有关拆分模型的摘要。

图片 12-12
拆分模型查看器

分割组

ed	准确度	模型大小（突触）	记录
Did not complete high school	68.3%	62	372
High school degree	66.8%	42	198
Some college	72.3%	42	87
College degree	59.2%	12	38
Post-undergraduate degree	.	.	.

不能为一个或多个分割组构件模型。

拆分。列标题显示用于创建拆分的字段，而单元格则是拆分值。双击任意拆分打开“模型查看器”以用于该拆分的建模。

精确性。百分比形式的整体精确性。

模型大小。模型大小取决于建模方法：对于树模型，它是树中节点的数量；对于线性模型，它是系数的数量；对于神经网络，它是神经元的数量。

记录。训练样本中输入记录的加权数量。

章

13

使用命令语法

这种功能强大的命令语言，使您可以保存和自动执行许多常见任务。它还提供一些在菜单和对话框中没有的功能。

大多数命令可以从菜单和对话框访问。但是，某些命令和选项只能通过命令语言使用。命令语言还允许您将作业保存在语法文件中，以便将来可以重复进行分析，或者通过生产作业在自动化作业中运行该语法文件。

语法文件就是一个包含命令的文本文件。虽然可以打开语法窗口并键入命令，但如果您使用以下一种方法，让软件帮助您构建语法文件，那么事情就会简单很多：

- 从对话框粘贴命令语法
- 从输出日志复制语法
- 从日志文件复制语法

命令语法的详细参考信息，以两种方式提供：集成在完整的“帮助”系统中，在“帮助”菜单的命令语法参考中，以单独的 PDF 格式文档提供。按 F1 键可获得语法窗口中当前命令的上下文相关的帮助。

语法规则

当您在会话过程中从命令语法窗口运行命令时，将以**交互**模式运行命令。

以下规则适用于交互模式下的命令规范：

- 每个命令必须从新行开始。命令可在命令行的任何列中开始，并持续所需任意数量的行。END DATA 命令是个例外，该命令必须在数据结尾之后第一行的第一列中开始。
- 每个命令应该以句点为命令终止符。但是，最好省略 BEGIN DATA 中的终止符，以便将内联数据视为一个连续指定。
- 命令终止符必须是命令中的最后一个非空白字符。
- 如果没有句点作为命令终止符，则将空行解释为命令终止符。

注意：为与命令的其他执行模式（包括在交互式会话中用 INSERT 或 INCLUDE 命令运行的命令文件）兼容，命令语法的每行不能超过 256 个字节。

- 大多数子命令由斜杠 (/) 分隔。第一条子命令前面的斜杠通常是可选的。
- 变量名必须使用全名拼写。
- 单引号或双引号内包含的文本必须包含在单独一行上。
- 区域设置或语言环境设置如何，必须使用句点 (.) 指示小数。
- 以句点结束的变量名可能在由对话框创建的命令中导致错误。不能在对话框中创建这样的变量名，并且通常应避免这样的变量名。

命令语法不区分大小写，三个字母或四个字母的缩写可用于许多命令规范。可以使用任意多行指定单独一条命令。可以在几乎所有允许单个空格的地方添加空格或换行符，例如斜杠、括号和算术运算符两端或变量名之间。例如，

```
FREQUENCIES
 VARIABLES=JOBCAT GENDER
 /PERCENTILES=25 50 75
 /BARCHART.
```

和

```
freq var=jobcat gender /percent=25 50 75 /bar.
```

都是可接受的替代形式，它们生成相同的结果。

INCLUDE 文件

对于通过 INCLUDE 命令运行的命令文件，可使用批处理模式语法规则。

以下规则适用于批处理模式下的命令规范：

- 命令文件中的所有命令必须以列 1 开始。如果要缩进指定的命令以提高命令文件的可读性，则可以在第一列中使用加号 (+) 或减号 (-)。
- 如果对命令使用多行，则每个连续行的列 1 必须为空。
- 命令终止符是可选的。
- 一行不能超过 256 个字节；将截断任何额外的字符。

除非现有的命令文件已使用了 INCLUDE 命令，否则应改用 INSERT 命令，因为它适用于符合任意规则的命令文件。如果通过将对话框选择粘贴到语法窗口中来生成命令语法，则这些命令的格式适用于任何操作模式。有关更多信息，请参见命令语法参考（可从“帮助”菜单获得 PDF 格式的文件）。

从对话框粘贴语法

建立命令语法文件的最简单方法是在对话框中进行选择，然后将选择的语法粘贴到语法窗口中。通过在冗长分析的每一步粘贴语法，您可以构建作业文件，该文件使您可以在将来重复进行分析，或者通过生产工具运行自动化作业。

在语法窗口中，您可以运行粘贴的语法，对其进行编辑，并将其保存到语法文件。

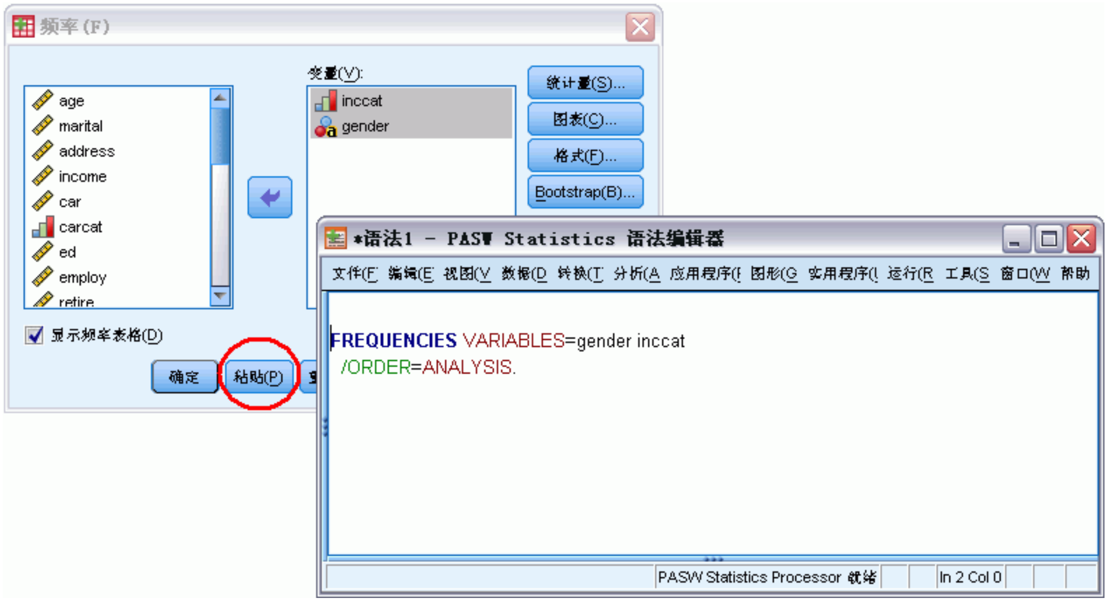
从对话框粘贴语法

▶ 打开对话框并选择所需的语法。

▶ 单击**粘贴**。

命令语法会粘贴到指定的语法窗口中。如果没有打开语法窗口，则会自动打开新的语法窗口，语法会粘贴在那里。默认情况下，语法会粘贴到最后一条命令的后面。您可以选择将语法粘贴到光标位置，或者覆盖选中的语法。该设置在“选项”对话框的“语法编辑器”选项卡上指定。

图片 13-1
从对话框粘贴的命令语法

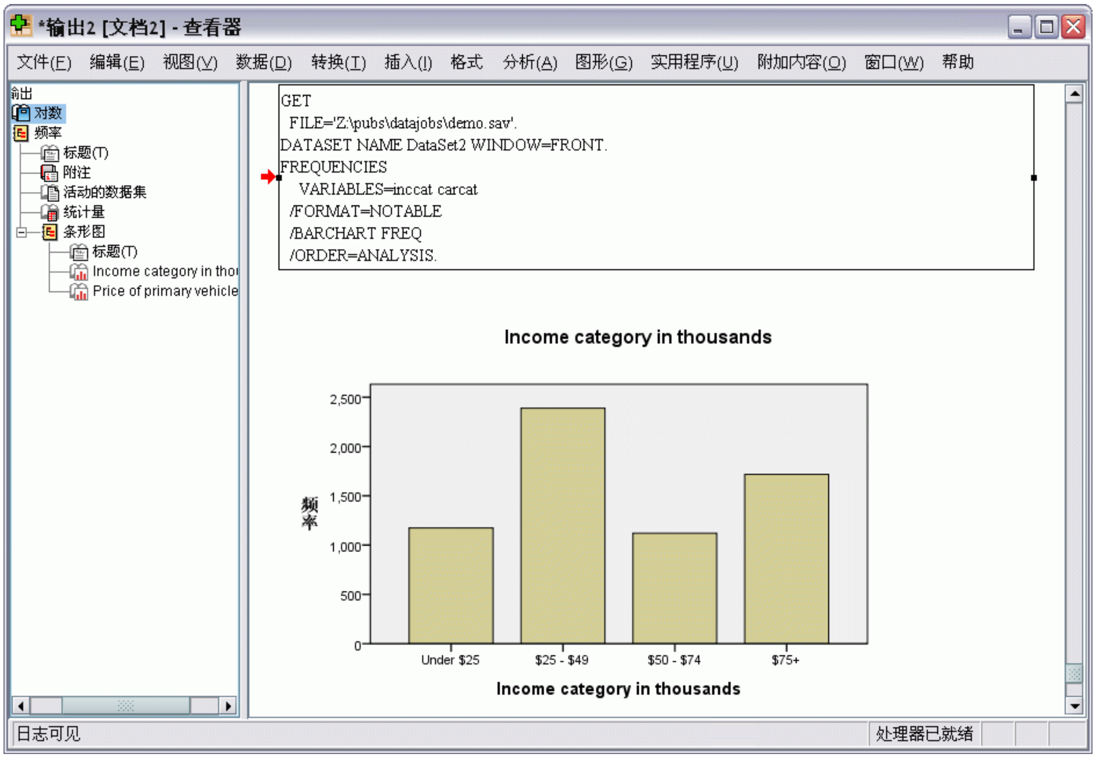


从输出日志复制语法

您可以通过从显示在查看器中的日志复制命令语法来建立语法文件。要使用此方法，必须在运行分析之前，在“查看器”设置（“编辑”菜单，“选项”，“查看器”选项卡）中选择**在日志中显示命令**。然后每条命令将随同分析的输出一起显示在查看器中。

在语法窗口中，您可以运行粘贴的语法，对其进行编辑，并将其保存到语法文件。

图片 13-2
日志中的命令语法



从输出日志复制语法

▶ 在运行分析之前，从菜单中选择：
编辑 > 选项...

▶ 在“查看器”选项卡上，选择**在日志中显示命令**。

运行分析的同时，在对话框中选择的命令将记录到日志中。

▶ 打开先前保存的语法文件或新建一个语法文件。要新建语法文件，从菜单中选择：
文件 > 新建 > 语法

▶ 在查看器中，双击日志项以将其激活。

▶ 选择要复制的文本。

▶ 从“查看器”菜单中选择：
编辑 > 粘贴

▶ 在语法窗口中，从菜单中选择：
编辑 > 粘贴

使用语法编辑器

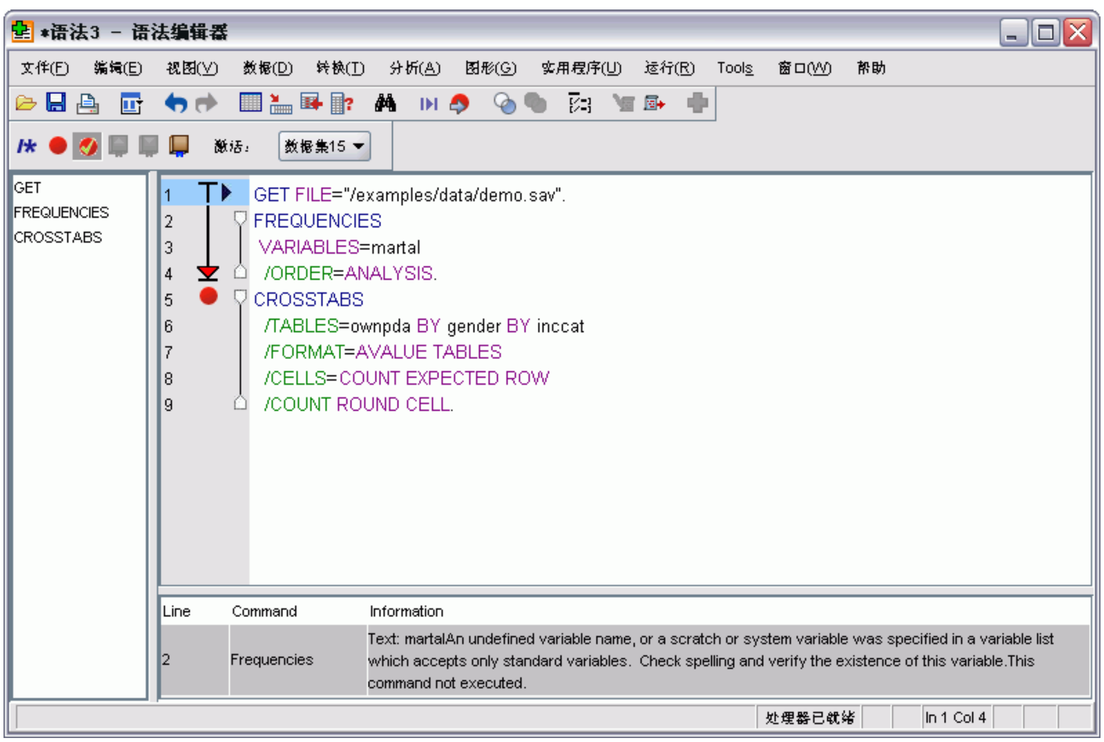
语法编辑器提供一个专为创建、编辑和运行命令语法而设计的环境。语法编辑器有以下特色：

- **自动完成。** 随着您的输入，您可以从上下文敏感列表中选择命令、子命令、关键字和关键字值。您可以选择自动提示列表或按需要显示列表。
- **颜色编码。** 命令语法（命令、子命令、关键字和关键字值）的识别元素是颜色编码，因此浏览一下即可定位未识别项。另外，一些常见语法错误—如未匹配的引号—都经过颜色编码以供快速识别。
- **分界点。** 您可以在指定点停止执行命令语法，这允许您在查看数据或输出后再继续。
- **书签。** 您可以设置书签，以允许您快速导航大型命令语法文件。
- **自动缩进。** 您可以使用与从对话框粘贴的语法类似的缩进样式来自动设置您的语法格式。
- **逐步执行。** 您可以一次一个命令地逐步执行命令语法，使用单击前进到下一个命令。

注意：当使用从右至左语言时，建议在“选项”对话框中的“语法编辑器”选项卡上选中“针对从右至左语言优化”复选框。

语法编辑器窗口

图片 13-3
语法编辑器



语法编辑器窗口分为四个区域：

- 编辑器窗格是语法编辑器窗口的主要部分，是您输入和编辑命令语法的地方。
- 装订线与编辑器窗格相邻，显示行号和分界点位置等信息。
- 导航窗格位于装订线和编辑器窗格左侧，在语法编辑器窗口中显示所有命令列表，单击即可导航到任何命令。
- 错误窗格位于编辑器窗格下方，显示运行时间错误。

装订线内容

行号、分界点、书签、命令跨度和进度指示显示在语法窗口中编辑窗格的左侧。

- 行号不考虑 INSERT 和 INCLUDE 命令中所引用的任何外部文件。您可以通过从菜单中选择“视图 > 显示行号”以显示或隐藏行号。
- 分界点在指定点停止执行，表示为一个与设置分界点的命令相邻的红圈。
- 书签在命令语法文件中标记特定行，表示为包含分配到书签的数字（1-9）的正方形。悬停在书签图标上将显示分配到书签的书签编号以及名称（如果有）。
- 命令跨度是提供命令开始和结束的可视指示符的图标。您可以通过从菜单中选择“视图 > 显示命令跨度”以显示或隐藏命令跨度。
- 给定语法运行的进度在装订线中使用向下箭头表示，从第一个命 运行扩展到最后一个命令运行。这在运行包含分界点的命令语法和逐步执行命令语法时最有用。

导航窗格

导航窗格在语法窗口中包含所有已识别命令列表，以它们在窗口中发生的顺序显示。单击导航窗格中的命令会将光标置于命令开始。

- 您可以使用向上和向下箭头键移动通过命令列表或单击命令以导航到该命令。双击将选择命令。
- 包含某些类型的语法错误—如未匹配的引号—的命令名称默认情况下显示为红色加粗文本。
- 未识别文本的每行第一个单词显示为灰色。
- 您可以通过从菜单中选择“视图 > 显示导航窗格”以显示或隐藏导航窗格。

错误窗格

错误窗格显示最近运行中的运行时间错误。

- 每个错误的信息包含存在错误的命令的起始行号。
- 您可以使用向上和向下箭头移动通过错误列表。
- 单击列表中的一个条目会将光标置于生成该错误的命令的第一行上。
- 您可以通过从菜单中选择“视图 > 显示错误窗格”以显示或隐藏错误窗格。

使用多个视图

您可以将编辑器窗格拆分为上下两个窗格。

▶ 从菜单中选择：
窗口 > 拆分

在导航和错误窗格中的操作，例如单击某个错误，将对光标所在的窗格起作用。

您可以双击窗格或选择“窗口 > 删除拆分”以删除窗格拆分。

术语

命令。 语法的基本单位是命令。每个命令以命令名称开头，命令名称包括一个、两个或三个单词—例如，DESCRIPTIVES、SORT CASES 或 ADD VALUE LABELS。

子命令。 大多数命令包含子命令。子命令提供附加规范，以正斜杠开头，后面跟子命令名称。

关键字。 关键字是通常在子命令中使用的固定术语，用来指定子命令的可用选项。

关键字值。 关键字可以拥有如指定一个选项或数字值的固定术语之类的值。

示例

```
CODEBOOK gender jobcat salary
 /VARINFO VALUELABELS MISSING
 /OPTIONS VARORDER=MEASURE.
```

- 命令名称是 CODEBOOK。
- VARINFO 和 OPTIONS 是子命令。
- VALUELABELS、MISSING 和 VARORDER 是关键字。
- MEASURE 是与 VARORDER 相关的关键字值。

自动完成

语法编辑器以自动完成命令、子命令、关键字和关键字值的形式提供帮助。默认情况下，它将在您输入时提示一个可用术语的上下文敏感列表。按 Enter 或 Tab 键将在光标位置插入列表中当前突出显示的项目。您可以通过按 Ctrl+空格按需要显示列表，也可以通过按 Esc 键关闭列表。

“工具”菜单上的“自动完成”菜单项可以在自动显示或不自动显示自动完成列表之间进行切换。您还可从“选项”对话框中的“语法编辑器”选项卡启用或禁用自动显示列表。切换自动完成菜单项会覆盖“选项”对话框上的设置，但不会在会话之间保持不变。

注意：如果输入一个空格，自动完成列表将关闭。对于由多个单词组成的命令—如 ADD FILES—选择所需命令，然后输入任何空格。

颜色编码

语法编辑器颜色代码识别命令语法元素，如命令和子命令，以及各种语法错误，如未匹配的引号或圆括号。未识别的文本未经颜色编码。

命令。 默认情况下，识别的命令显示为蓝色粗体文本。但是，如果在命令中有一个识别的语法错误—如缺失的圆括号—命令名称默认情况下显示为红色加粗文本。

注意：命令名称的简写—如 FREQUENCIES 简写为 FREQ —不标为彩色，但此简写有效。

子命令。 别的子命令默认情况下显示为绿色。但是，如果子命令缺失一个所需等号或随后跟的是一个无效等号，子命令名称默认情况下显示为红色。

关键字。 识别的关键字默认情况下显示为褐紫红色。但是，如果关键字缺失一个所需等号或随后跟的是一个无效等号，关键字默认情况下显示为红色。

关键字值。 识别的关键字值默认情况下显示为橙色。用户指定的关键字值未经颜色编码，如整数、实数和引号引起的字符串。

注释。 注释中的文本默认情况下显示为灰色。

引号。 引号及其中的文本默认情况下显示为黑色。

语法错误。 与以下语法错误相关的文本默认情况下显示为红色。

- **未匹配的圆括号、方括号和引号。** 未检测到注释和引号引起的字符串中未匹配的圆括号和方括号。在引号引起的字符串中未匹配的单引号或双引号在语法上无效。

  某些命令包含不是命令语法的文本块—如 BEGIN DATA-END DATA、BEGIN GPL-END GPL 和 BEGIN PROGRAM-END PROGRAM。在这种块中未检测到未匹配的值。
- **长行。** 长行是包含 251 个以上字符的行。
- **结束语句。** 一些命令在命令终止符之前需要 END 语句（例如，BEGIN DATA-END DATA）或在命令流中的稍后某个时刻需要匹配的 END 命令（例如，LOOP-END LOOP）。在这两种情况下，命令默认情况下显示为红色，直到添加了所需 END 语句。

注意：通过从“工具”菜单的“验证错误”子菜单中选择“下一个错误”或“上一个错误”，您可以导航到下一个或上一个语法错误。

从“选项”对话框中的“语法编辑器”选项卡，您可以更改默认颜色和文本样式，也可以关闭或打开颜色编码。您还可通过从菜单中选择“工具 > 颜色编码”关闭或打开命令、子命令、关键字和关键字值的颜色编码。您可以通过选择“工具 > 验证”关闭或打开语法错误的颜色编码。在“工具”菜单上进行的选择会覆盖“选项”对话框中的设置，但不会在会话之间保持不变。

注意：不支持“宏”中的命令语法颜色编码。

分界点

分界点允许您在语法窗口中的指定点停止执行命令语法，并在准备好时继续执行。

- 分界点设置在命令级别，并在运行命令之前停止执行。
- 分界点无法在 LOOP-END LOOP、DO IF-END IF、DO REPEAT-END REPEAT、 INPUT PROGRAM-END INPUT PROGRAM 和 MATRIX-END MATRIX 块中发生。但是，它们可以设置在这种块的开始，并在运行块之前停止执行。
- 分界点无法设置在包含非 IBM® SPSS® Statistics 命令语法的行上，如在 BEGIN PROGRAM-END PROGRAM、BEGIN DATA-END DATA 和 BEGIN GPL-END GPL 块中发生。

- 分界点不与命令语法文件一起保存，不包括在复制的文本中。
- 默认情况下，在执行过程中接受分界点。您可以从“工具 > 接受分界点”中切换是否接受分界点。

插入分界点

▶ 单击命令文本左侧装订线的任意位置。

或

▶ 将光标置于所需命令中。

▶ 从菜单中选择：
工具 > 切换断点

分界点在命令文本左侧装订线中以及与命令名称相同的行上表示为红圈。

清除分界点

清除单个分界点：

▶ 在命令文本左侧装订线中单击代表分界点的图标。

或

▶ 将光标置于所需命令中。

▶ 从菜单中选择：
工具 > 切换断点

清除所有分界点：

▶ 从菜单中选择：
工具 > 清除所有断点

参见 运行命令语法第 260 页码 以获取有关在存在分界点的情况下的运行时间行为的信息。

书签

书签允许您快速导航到命令语法文件中的指定位置。您在给定文件中可以拥有最多 9 个书签。书签与文件一起保存，但在复制文本时不包括。

插入书签

▶ 将光标定位到想要插入书签的行上。

▶ 从菜单中选择：
工具 > 切换书签

新书签被分配从 1 到 9 的下一个可用数字。表示为括起分配编号的方括号，并显示在命令文本左侧装订线中。

清除书签

清除单个书签：

- ▶ 将光标置于包含书签的行上。
- ▶ 从菜单中选择：
工具 > 切换书签

清除所有书签：

- ▶ 从菜单中选择：
工具 > 清除所有书签

重命名书签

您可以将一个名称与一个书签关联起来。这是除在创建时分配给书签的数字（1-9）以外的另一名称。

- ▶ 从菜单中选择：
工具 > 重新命名书签
- ▶ 为书签输入名称并单击**确定**。

指定的名称将替换书签的任何现有名称。

导航书签

导航到下一个或上一个书签：

- ▶ 从菜单中选择：
工具 > 下一个书签
或
工具 > 上一个书签

导航到特定书签：

- ▶ 从菜单中选择：
工具 > 转至书签
- ▶ 选择所需的书签。

注释或取消注释文本

您可以注释整个命令以及未识别为命令语法的文本，还可以取消注释先前被注释的文本。

注释文本

- ▶ 选择所需的文本。注意，如果选择了命令的任何部分，将对命令进行注释。
- ▶ 从菜单中选择：
工具 > 切换注释选择

您可以通过将光标置于命令中的任意位置并选择“工具 > 切换注释选择”来注释单个命令。

取消注释文本

- ▶ 选择要取消注释的文本。注意，如果选择了命令的任何部分，将对命令取消注释。
- ▶ 从菜单中选择：

工具 > 切换注释选择

您可以通过将光标置于命令中的任意位置并选择“工具 > 切换注释选择”来取消注释单个命令。注意，此功能不会从命令中删除注释（以 /* 和 */ 开始的文本）或通过 COMMENT 关键字创建的注释。

设置语法格式

您可以增加或减少选定语法行的缩进，还可以自动调整选定内容的缩进，以使语法格式与从对话框粘贴的语法格式类似。

- 默认缩进为四个空格，应用于选定语法行的缩进，以及自动缩进。您可以在“选项”对话框的“语法编辑器”选项卡上更改缩进大小。
- 注意，在“语法编辑器”中使用 Tab 键不会插入制表符。它会插入一个空格。

缩进文本

- ▶ 选择所需文本或将光标定位到您要缩进的单行上。
- ▶ 从菜单中选择：

工具 > 语法缩进 > 增加缩进

您还可以按 Tab 键缩进选定内容或行。

减少缩进

- ▶ 选择所需文本或将光标定位到您要减少缩进的单行上。
- ▶ 从菜单中选择：

工具 > 语法缩进 > 减少缩进

自动缩进文本

- ▶ 选择所需的文本。
- ▶ 从菜单中选择：

工具 > 语法缩进 > 自动缩进

在您自动缩进文本时，任何现有缩进将被删除，并替换为自动产生的缩进。注意，在 BEGIN PROGRAM 块中的自动缩进代码可能会破坏那些依赖于特定缩进而工作的代码，例如包含循环和条件块的 Python 代码。

采用自动缩进功能的格式语法不能在批处理模式下运行。例如，自动缩进 INPUT PROGRAM-END INPUT PROGRAM、LOOP-END LOOP、DO IF-END IF 或 DO REPEAT-END REPEAT 块，将导致在批处理模式中语法失败，因为块中的命令将被缩进，而不会从列 1 开始，这与批处理模式的要求不符。不过，您可以在批处理模式下使用 -i 开关，强制 Batch Facility 使用交互语法规则。

运行命令语法

▶ 在语法窗口突出显示想要运行的命令。

▶ 单击语法编辑器工具栏上的“运行”按钮（指向右边的三角形）。运行所选命令，或者如果没有选择，运行光标所在命令。

或

▶ 从“运行”菜单中选择一项。

- **全部。**运行语法窗口中的所有命令，接受任何分界点。
- **选择。**运行当前选择的命令，接受任何分界点。这包括任何只有部分突出显示的命令。如果没有选择，则运行光标所在的命令。
- **至结束。** 运行从当前选择中的第一个命令到语法窗口中的最后一个命令的所有命令，接受任何分界点。如果没有选择，则从光标所在的命令开始运行。
- **逐步执行。** 一次一个命令地从语法窗口中的第一个命令（从开始逐步执行）或从当前选择中的光标所在的命令（从当前逐步执行）运行命令语法。如有选择的文本，则从选择中的第一个命令开始运行。在给定命令运行之后，光标前进到下一个命令，您可以通过选择“继续”继续逐步执行。

  LOOP-END LOOP、DO IF-END IF、DO REPEAT-END REPEAT、INPUT PROGRAM-END INPUT PROGRAM 和 MATRIX-END MATRIX 块在使用“逐步执行”时被视为单个命令。您无法步进到其中任何一个块。
- **继续。** 继续被分界点或逐步执行停止的运行。

进度指示

给定语法运行的进度在装订线中使用向下箭头表示，跨到最后一组命令运行。例如，您选择运行包含分界点的语法窗口中的所有命令。在第一个分界点，箭头将从窗口中的第一个命令跨到包含该分界点之前的命令。在第二个分界点，箭头将从包含第一个分界点的命令扩展到包含第二个分界点之前的命令。

带有分界点的运行时间行为

- 当运行包含分界点的命令语法时，在每个分界点执行停止。具体而言，从给定分界点（或运行开始）到下一个分界点（或运行结束）的命令语法块被提交以供执行，就像您已选择该语法并选择“运行 > 选择”一样。

- 您可以使用多个语法窗口，每个都有自己的一组分界点，但是执行命令语法的队列只有一个。一旦提交了一个命令语法块（如到第一个分界点的命令语法块），则不会执行其他命令语法块，直到前一个块已完成，无论块是否在相同或不同的语法窗口中。
- 执行在一个分界点处停止，您可以在其他语法窗口中运行命令语法，并检查数据编辑器或查看器窗口。但是，修改包含分界点的语法窗口的内容或更改光标在该窗口中的位置将取消运行。

Unicode 语法文件

在 Unicode 模式下，保存在会话期间创建或修改的命令语法文件的默认格式也为 Unicode (UTF-8)。IBM® SPSS® Statistics 16.0 之前的版本不能读取 Unicode 格式命令语法文件。有关 Unicode 模式的更多信息，请参见一般选项第 290 页码。

要以兼容于以前版本的格式保存语法文件：

▶ 从语法窗口菜单中选择：
文件 > 另存为

▶ 从“另存为”对话框的“编码”下拉列表中，选择**本地编码**。本地编码由当前区域决定。

多条执行命令

从对话框粘贴或者从日志或日志文件中复制的语法可能包含 EXECUTE 命令。从语法窗口运行命令时，通常不需要 EXECUTE 命令，并且该命令可能会降低性能，尤其是对于大型数据文件，因为每个 EXECUTE 命令都读取整个数据文件。有关更多信息，请参见命令语法参考（可从任意 IBM® SPSS® Statistics 窗口的“帮助”菜单进行查看）中的 EXECUTE 命令。

延迟函数

一个值得注意的例外情况是包含延迟函数的转换命令。在没有任何干扰性 EXECUTE 命令或者其他读取数据的命令的一系列转换命令中，延迟函数是在所有其他转换之后计算的，与命令顺序无关。例如，

```
COMPUTE lagvar=LAG(var1).
COMPUTE var1=var1*2.
```

和

```
COMPUTE lagvar=LAG(var1).
EXECUTE.
COMPUTE var1=var1*2.
```

对于 lagvar 的值生成不同的结果，因为前者使用转换后的 var1 值，而后者使用原始值。

章

14

图表工具的概述

通过“图形”菜单和“分析”菜单中的多个过程创建高分辨率图表和图。本章提供图表工具的概述。

生成和编辑图表

创建图表之前，您需要从数据编辑器中获取数据。您可以将数据直接输入数据编辑器，打开先前保存的数据文件，或读取一个电子表格、制表符分隔的数据文件或数据库文件。“帮助”菜单中的“教程”选择有创建和修改图表的联机示例，联机帮助系统提供有关创建和修改所有图表类型的信息。

生成图表

使用图表生成器可以根据预定义的图库图表或图表的单独部分（例如，轴和条形）生成图表。可以通过将图库图表或基本元素拖放到**画布**（“图表生成器”对话框上“变量”列表右侧的较大区域）上来生成图表。

生成图表时，画布会显示图表的预览。虽然预览使用已定义的变量标签和测量级别，但预览并不会显示实际的数据。它使用随机生成的数据简略勾勒出图表的外观。

对于新用户而言，使用图库是首选方法。有关使用图库的信息，请参见根据图库生成图表第 263 页码。

如何启动图表生成器

▶ 从菜单中选择：

图形 > 图表生成器

此操作将打开“图表生成器”对话框。

图片 14-1
“图表生成器”对话框



根据图库生成图表

生成图表最简单的方法是使用图库。以下是根据图库生成图表的常规步骤。

- ▶ 如果尚未显示**图库**选项卡，请单击它。
- ▶ 在“选择范围”列表中，选择一个图表类别。每个类别都提供了多种类型。
- ▶ 将所需图表的图片拖到画布上。也可以双击该图片。如果画布已显示了一个图表，则图库图表会替换该图表上的轴系和图形元素。
- ▶ 将变量从“变量”列表拖放至轴放置区以及分组放置区（如果有）。如果轴放置区已显示统计量，而您要使用该统计量，则不必将变量拖到放置区中。只有区域中的文本为蓝色时才需要向该区域添加变量。如果文本为黑色，则表明该区域已包含变量或统计量。

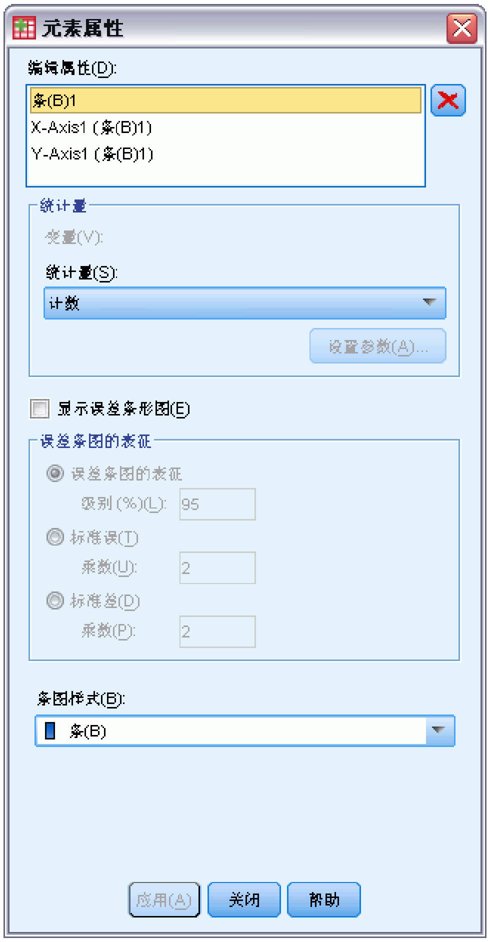
注意：变量的测量级别很重要。生成图表时，图表生成器根据测量级别设置缺省值。此外，测量级别不同，产生的图表看上去也会不同。可以临时更改变量的测量级别，方法是右键单击该变量，然后选择选项。

图片 14-2
带有完整拖放区域的“图表生成器”对话框



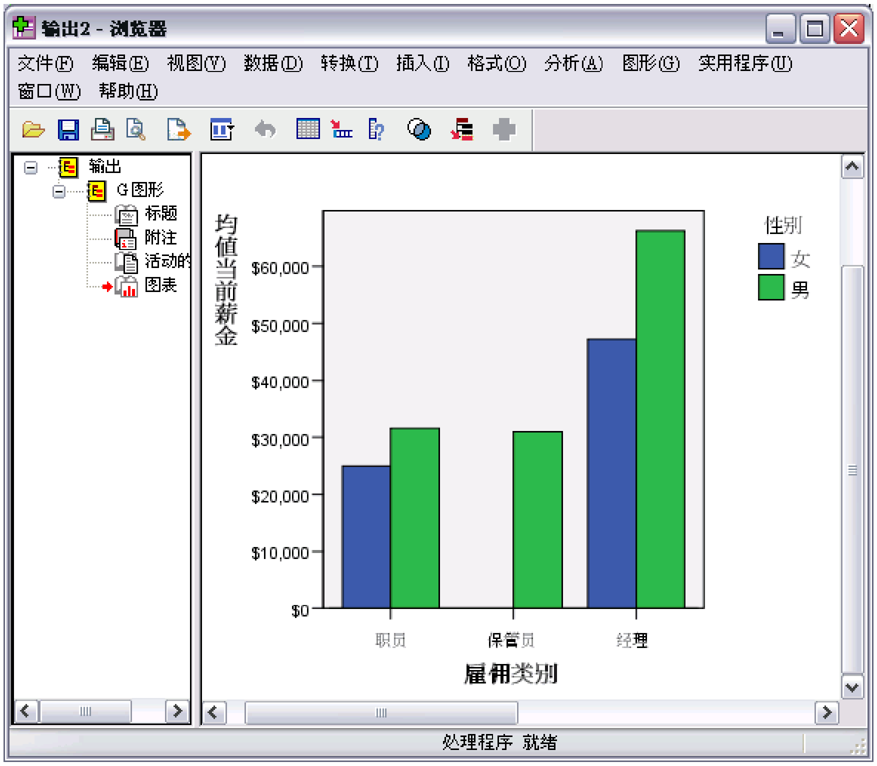
▶ 如果需要更改统计量或修改轴或图注的属性（如刻度范围），请单击**元素属性**。

图片 14-3
“元素属性”对话框



- 在“编辑属性”列表中，选择要更改的项。（有关特定属性的信息，请单击**帮助**。）
- 进行任何更改后，请单击**应用**。
- 如果需要将更多变量添加到图表中（例如要进行分类汇总或将变量放入面板），可在“图表生成器”对话框中单击**组/点 ID** 选项卡并选择一个或多个选项。然后，将分类变量拖放至画布上显示的新放置区。
- 如果要转置图表（例如，使条变为水平的），请单击**基本元素**选项卡，然后单击**转置**。
- 单击**确定**以创建图表。图表将显示在浏览器中。

图片 14-4
“浏览器”窗口中显示的条形图



编辑图表

图表编辑器提供功能强大、易于使用的环境，供您定制图表和探索数据。图表编辑器有以下特色：

- **简单、直观的用户界面。**使用菜单、上下文菜单和工具栏可以很快地选择并编辑部分图表。也还可以直接在图表上输入文本。
- **众多格式和统计选项。**您可以选择各种风格和统计选项。
- **强大的探索工具。**您可以用各种方式探索数据，如标记数据，对数据进行重新排序，或者旋转数据。可以更改图表类型和图表中的变量角色。还可以添加分布曲线、拟合线、内插线和参考线。
- **灵活的模板，使外观和操作保持一致。**可以创建定制的模板，并用这些模板轻松创建具有所希望的外观和选项的图表。例如，如果希望坐标轴标签始终朝向特定方向，就可以在模板中指定这个方向，然后将模板应用于其他图表。

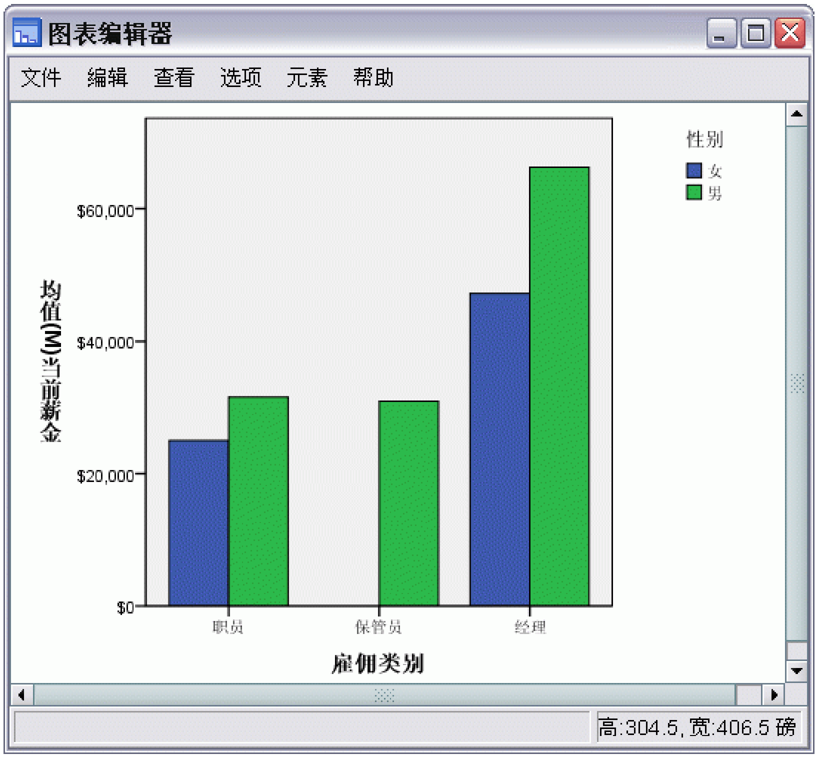
如何查看图表编辑器

▶ 在 IBM® SPSS® Statistics 中创建图表，或打开包含图表的浏览器文件。

▶ 双击浏览器中的一个图表。

图表将显示在图表编辑器中。

图片 14-5
在图表编辑器中显示图表



图表编辑器基础

图表编辑器提供多种操作图表的方法。

菜单

图表编辑器中可以执行的很多操作都是用菜单完成的，特别是将某项添加到图表的时候。例如，使用菜单将拟合线添加到散点图。将某项添加到图表后，经常需要使用“属性”对话框来指定对应所添加的项的选项。

“属性”对话框

在“属性”对话框可以找到对应图表及其图表元素的选项。

要查看“属性”对话框，您可以：

▶ 双击图表元素。

或

▶ 选择图表元素，然后在菜单中选择：
编辑 > 属性

此外，“属性”对话框在您将某项添加到图表时自动出现。

图片 14-6
“属性”对话框，“填充”和“边框”选项卡



使用“属性”对话框的选项卡可以设置选项，并对图表作出其他更改。在“属性”对话框显示的选项卡取决于您当前的选择。

有些选项卡包含预览，可使您了解应用所做的更改后对当前选择的影响。但是图表本身并不反映您所作的更改，只有单击**应用**才会反映出来。可以在多个选项卡上作出更改，然后单击**应用**。如果您必须更改选择以修改图表上不同的元素，则在更改选择前单击**应用**。如果没有在更改选择前单击**应用**，则以后单击**应用**只能将更改应用于当前选定的元素。

只有特定设置可用，这取决于您的选择。各个选项卡的帮助指定查看选项卡需要选择的内容。如果选择了多个元素，则只能更改所有元素共有的设置。

工具栏

工具栏提供“属性”对话框中某些功能的快捷方式。例如，可以不用“属性”对话框的“文本”选项卡，而使用“编辑”工具栏来更改文本的字体和样式。

保存更改

关闭图表编辑器时，将保存图表修改。修改过的图表将随后显示在浏览器中。

图表定义选项

在图表生成器中定义图表时，可以为图表创建添加标题和更改选项。

添加和编辑标题和脚注

可以向图表添加标题和脚注，以帮助浏览器解释图表。图表生成器还会自动在脚注中显示误差条图信息。

如何添加标题和脚注

- 单击**标题/脚注**选项卡。
- 选择一个或多个标题和脚注。画布会显示一些文本，用于指示这些标题和脚注已添加到图表中。
- 使用“元素属性”对话框可编辑标题/脚注文本。

如何移去标题或脚注

- 单击**标题/脚注**选项卡。
- 取消选择要移去的标题或脚注。

如何编辑标题或脚注文本

在添加标题和脚注时，不能直接在图表上编辑与其关联的文本。图表生成器中的其他项相同，可以使用“元素属性”对话框进行编辑。

- 如果未显示“元素属性”对话框，则单击**元素属性**。
- 在“编辑属性”列表中，选择标题、子标题或脚注（如**标题 1**）。
- 在“内容”框中，键入与标题、子标题或脚注关联的文本。
- 单击**应用**。

设置一般选项

图表生成器为图表提供一般选项。这些选项适用于整个图表，而不是图表上的特定项。一般选项包含缺失值处理、模板、图表大小和面板换行。

- 单击**选项**。
- 修改一般选项。随后是有关这些选项的详细信息。
- 单击**应用**。

用户缺失值

中断变量。 如果用于定义类别或子组的变量有缺失值，则选择**包含**，以便在图表中包含用户缺失值（由用户标识为缺失的值）的类别。这些“缺失”类别还在计算统计量时作为中断变量。“缺失”类别显示在类别轴上或图例中，例如，添加额外的条，或向饼图添加分区。如果没有缺失值，则不显示“缺失”类别。

如果选择此选项，并要在绘制完图表后取消显示，请在图表编辑器中打开图表，从“编辑”菜单中选择**属性**。使用“类别”选项卡将要取消显示的类别移到“排除”列表中。但是，请注意，如果隐藏“缺失”类别，则不会重新计算统计量。因此，像百分比统计量一类的内容仍将考虑“缺失”类别。

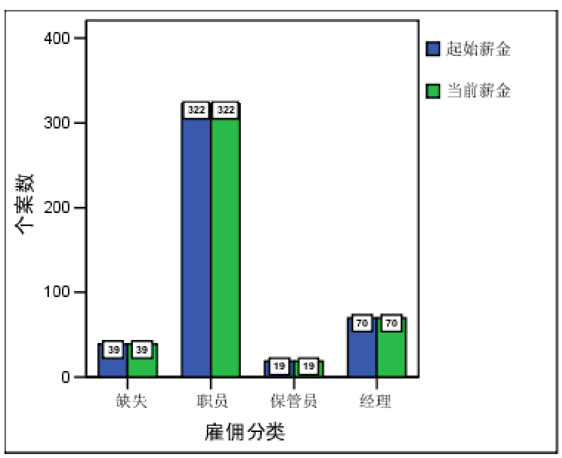
注意：此控制不影响系统缺失值。始终从图表中排除这些值。

摘要统计量和个案值。 可以选择以下选项之一，以排除具有缺失值的个案：

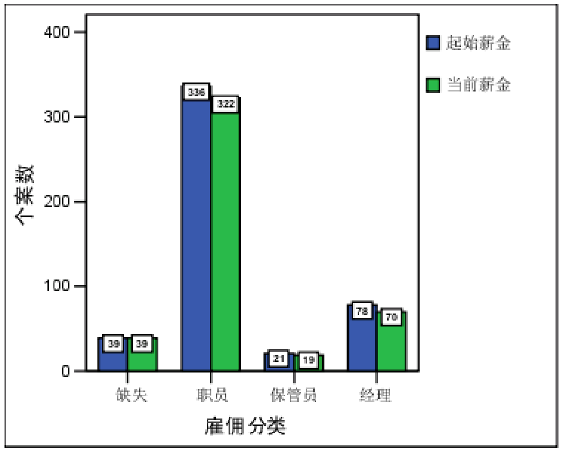
- **按列表排除以获取图表的一致基数。**如果图表中的任何变量对给定个案有缺失值，则从图表中排除整个个案。
- **逐个变量排除以最大限度地使用数据。**如果所选变量有缺失值，则在分析该变量时排除含有这些缺失值的个案。

要查看按列表排除缺失值和按逐个变量排除缺失值的差异，请考虑以下图片，它们为这两个选项的每一个都显示了一个条形图。

图片 14-7
按列表排除缺失值



图片 14-8
按逐个变量排除缺失值



数据包含当前薪金和雇佣类别变量中的一些系统缺失（空白）值。在其他一些个案中，输入值 0 并将之定义为缺失值。在两个图表中都选中**显示根据缺失值定义的组**选项，这将向其他显示的工作类别添加缺失类别。每一个图表的条形图标签中都将显示摘要函数值个案数。

在这两个图表中，26 个个案有工作类别的系统缺失值，13 个个案有用户缺失值 (0)。在按列表排除个案的图表中，因为无论何时一个值缺失，都会为所有变量排除该个案，所以每一条聚类中两个变量的个案数是等同的。在按逐个变量排除个案的图表中，不考虑其他变量中的缺失值，为类别中的每一变量绘制非缺失个案数。

模板

使用图表模板可以将一个图表的属性应用于另一图表。在图表编辑器中打开图表时，可以将其保存为模板。然后可以应用该模板，方法是创建时指定该模板，或以后在图表编辑器中应用它。

缺省模板。这是由选项指定的模板。可以通过从数据编辑器的“编辑”菜单中选择**选项**，然后单击**图表**来访问这些模板。首先应用缺省模板，这意味着其他模板可以覆盖它。

模板文件。单击**添加**，使用标准的文件选择对话框指定一个或多个模板。以模板的显示顺序应用这些模板。因此，位于列表末尾的模板可以覆盖位于列表开头的模板。

图表大小和面板

图表大小。指定大于 100 的百分比可将图表放大，指定小于 100 的百分比可将图表缩小。此百分比相对于缺省的图表大小。

面板。有很多面板列时，选择**使面板换行**可允许面板跨行换行，而不是强制放入一个特定的行中。除非选中此选项，否则面板将缩小以强制放入一行中。

章

15

使用预测模型对数据评分

应用预测模型到数据集的过程被称为**评分**数据。IBM® SPSS® Statistics 提供了用来构建预测模型（如回归、聚类、树和神经网络模型）的过程。构建模型后，可以将模型规范保存在文件中，该文件包含重建模型所需的所有信息。然后，您可以使用该模型文件在其他数据集中生成预测得分。(注意： 一些过程生成模型 XML 文件，而一些过程则生成 ZIP 存档文件。)

示例。 公司直销部门使用测试邮件的结果，为其联系人数据库的其余部分指定倾向得分，他们使用各种人口统计学特征来标识最有可能做出响应和购买产品的联系人。

可以将评分视为数据转换。模型在内部表现为一组数值转换，这组转换将应用于一组给定字段/变量（模型中指定的预测变量）以获得预测结果。在此意义下，用给定模型对数据评分的过程与对一组数据应用任何函数（例如平方根函数）在实质上相同。

评分过程包含两个基本步骤：

- 构建模型并保存模型文件。使用数据集构建兴趣结果（通常被称为**目标**）已知的模型。例如，如果您希望构建可预测谁可能会响应直接邮寄活动的模型，则需要从已包含响应人和未响应人信息的数据集开始。例如，这可能是对一小组客户发送的测试邮件的结果或来自过去类似活动的响应信息。

  注意：对于一些模型类型，没有兴趣的目标结果。例如，聚类模型没有目标，而一些最近邻模型也没有目标。

- 应用该模型到其他数据集（其中兴趣结果未知）以获取预测结果。

下表列出支持将模型规范导出为模型文件的过程。“选项”列显示包含该过程的附加选项。

过程名称	命令名称	选项
判别	DISCRIMINANT	Statistics Base
线性回归	REGRESSION	Statistics Base
自动线性模型	LINEAR	Statistics Base
二阶聚类	TWOSTEP CLUSTER	Statistics Base
最近邻元素	KNN	Statistics Base
Cox 回归	COXREG	Advanced Statistics
广义线性模型	GENLIN	Advanced Statistics
广义估计方程	GENLIN	Advanced Statistics

过程名称	命令名称	选项
广义线性混合模型	GENLINMIXED	Advanced Statistics
复杂样本一般线性模型	CSGLM	复杂抽样
复杂样本 Logistic 回归	CSLOGISTIC	复杂抽样
复杂样本序数回归	CSORDINAL	复杂抽样
复杂样本 Cox 回归	CSCOXREG	复杂抽样
Logistic 回归	LOGISTIC REGRESSION	回归
多项 Logistic 回归	NOMREG	回归
决策树	TREE	决策树
多层感知器	Multilayer Perceptron	神经网络
径向基函数	Radial Basis Functions	神经网络
异常检测	DETECTANOMALY	数据准备
Naive Bayes	NAIVEBAYES	SPSS Statistics 服务器

评分向导

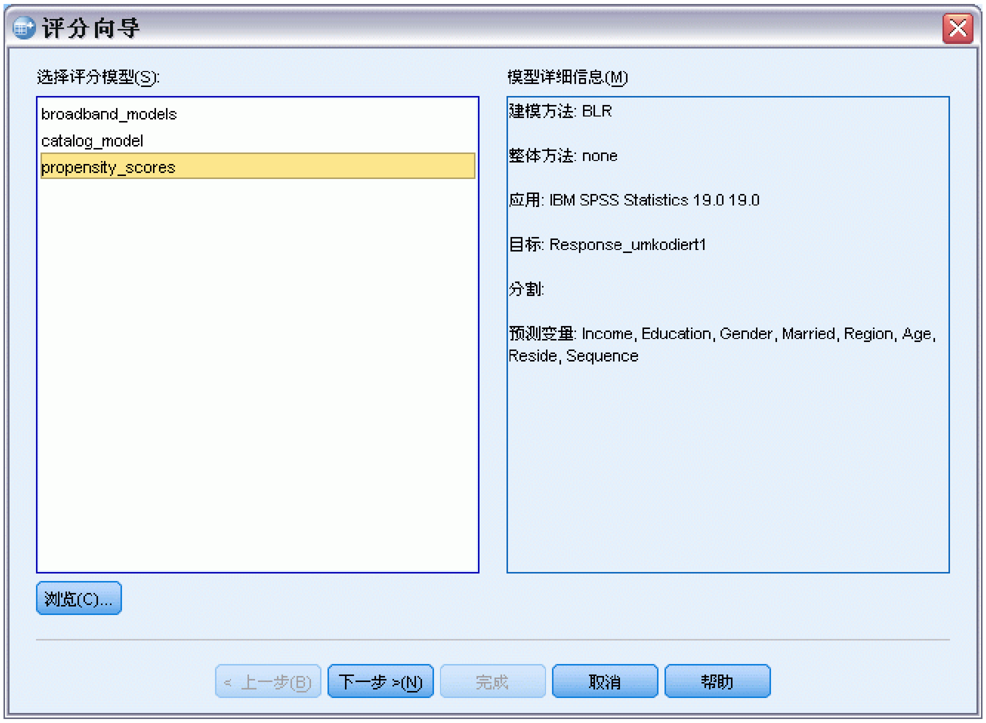
您可以使用评分向导将通过一个数据集创建的模型应用到另一个数据集，并生成**得分**，如兴趣结果的预测值和/或预测概率。

要使用预测模型对数据集进行评分

▶ 打开您要评分的数据集。

▶ 打开评分向导。从菜单中选择：
实用程序 > 评分向导。

图片 15-1
评分向导：选择评分模型



- ▶ 选择模型 XML 文件或 ZIP 存档。使用**浏览**按钮导航到其他位置以选择模型文件。
- ▶ 匹配存档数据集中的字段到模型中使用的字段。
- ▶ 选择您要使用的评分函数。

选择评分模型。 该模型文件可以是包含模型 PMML 的 XML 文件或 ZIP 存档。该列表只显示扩展名为 .zip 或 .xml 的文件；列表中不显示文件扩展名。您可以使用 IBM® SPSS® Statistics 创建的任何模型文件。您还可以使用由其他应用程序（如 IBM® SPSS® Modeler）创建的模型文件，但其中一些模型文件无法被 SPSS Statistics 读取，包括任何具有多个目标字段（变量）的模型。

模型详细信息。 这里显示所选模型的基本信息，如模型类型、目标（如果有）和用于构建模型的预测变量。由于必须读取模型文件才能获取该信息，因此在显示所选模型的此类信息之前可能会有延迟。如果 XML 文件或 ZIP 存档未被识别为 SPSS Statistics 可以读取的模型，则将显示信息提示您无法读取该文件。

匹配模型字段到数据集字段

为了对活动数据集进行评分，该数据集必须包含对应于模型中的所有预测变量的字段（变量）。如果模型还包含拆分字段，那么该数据集还必须包含对应于模型中所有拆分字段的字段。

- 默认情况下，自动匹配活动数据集中任何与模型中的字段具有相同名称和类型的字段。

- 使用下拉列表匹配数据集字段到模型字段。模型和数据集中每个字段的数据类型必须相同才能匹配字段。
- 在模型中的所有预测变量（以及拆分字段，如果有的话）与活动数据集中的字段匹配之前，您无法继续向导或对活动数据集进行评分。

图片 15-2
评分向导：匹配模型字段



数据集字段。 下拉列表包含活动数据集中所有字段的名称。无法选择不匹配相应模型字段数据类型的字段。

模型字段。 模型中使用的字段。

角色。显示的角色可以是以下角色之一：

- **预测变量。** 该字段在模型中用作预测变量。即，预测变量的值用于“预测”兴趣目标结果的值。
- **拆分。**拆分字段的值用于定义分组，其中每个分组单独评分。拆分字段值的每个唯一组合有一个单独分组。(注意： 拆分只对一些模型可用。)
- **记录 ID。** 记录（个案）标识。

度量。 模型中定义的字段的测量级别。对于测量级别会影响得分的模型，将使用模型中定义的测量级别，而不是活动数据集中定义的测量级别。有关测量级别的详细信息，请参见变量测量级别。

类型。 模型中定义的数据类型。活动数据集中的数据类型必须匹配模型中的数据类型。数据类型可以是以下类型之一：

- **字符串。** 活动数据集中，数据类型为字符串的字段匹配模型中的字符串数据类型。
- **数值。** 活动数据集中，显示格式不是日期或时间格式的数值字段匹配模型中的数值数据类型。其中包括 F（数值）、Dollar、Dot、Comma、E（科学计数法）和自定义货币格式。具有 Wkday（一周中的某天）和 Month（一年中的某月）格式的字段也被视为数值，而不是日期。对于一些模型类型，活动数据集中的日期和时间字段也被视为与模型中的数值数据类型匹配。
- **日期。** 活动数据集中，显示格式包含日期但不包含时间的数值字段匹配模型中的日期类型。其中包括 Date (dd-mm-yyyy)、Adate (mm/dd/yyyy)、Edate (dd.mm.yyyy)、Sdate (yyyy/mm/dd) 和 Jdate (dddyyyy)。
- **时间。** 活动数据集中，显示格式包含时间但不包含日期的数值字段匹配模型中的时间数据类型。其中包括 Time (hh:mm:ss) 和 Dtime (dd hh:mm:ss)
- **时间戳。** 活动数据集中，显示格式同时包含日期和时间的数值字段匹配模型中的时间戳数据类型。这对应于活动数据集中的 Datetime 格式 (dd-mm-yyyy hh:mm:ss)。

注意： 除了字段名和类型以外，您需要确保要评分的数据集中的实际数据值的记录方式与构建模型的数据集中的数据值相同。例如，如果模型使用 Income 字段构建，后者将收入划分为四种类别，而活动数据集中的 IncomeCategory 则将收入划分为六种类别或四种不同的类别，因此这些字段实际上彼此并不匹配，结果得分将不可靠。

缺失值

此组选项控制评分过程中遇到的模型中定义的预测变量缺失值的处理。评分过程中的缺失值指下列值之一：

- 预测变量不包含值。对于数值字段（变量），预测变量表示系统缺失值。对于字符串字段，预测变量表示空字符串。
- 在模型中，已将给定预测变量的值定义为用户缺失值。在活动数据集中（而未在模型中）定义为用户缺失的值在得分过程中并不被视为缺失值。
- 预测变量是分类变量，且其值不是模型中所定义的某个类别。

使用值替换。对具有缺失值的个案评分时尝试使用值替换。确定替换缺失值的值的方法取决于预测模型的类型。

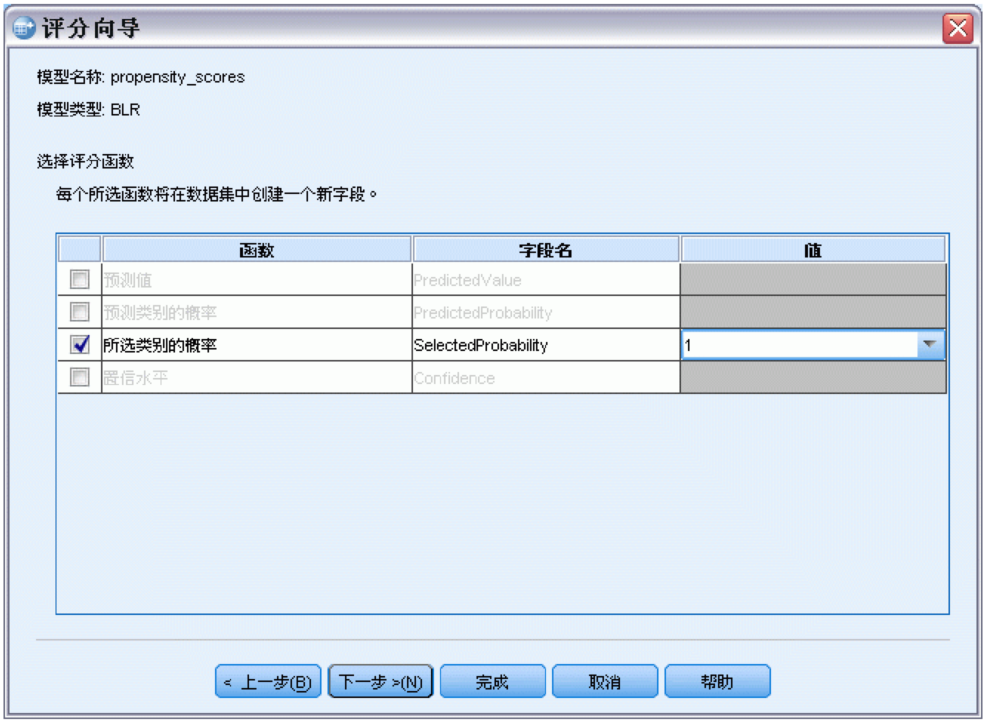
- **线性回归和判别模型。**对于线性回归和判别模型中的自变量，如果构建和保存模型时指定了缺失值的均值替换，则在评分计算中将使用此均值代替缺失值，评分过程继续。如果均值不可用，则返回系统缺失值。
- **决策树模型。** 对于 CHAID 和穷举 CHAID 模型，将为缺失的拆分变量选择最大的子节点。最大的子节点是使用学习样本个案的子节点中具有最大总体的子节点。对于 C&RT 和 QUEST 模型，首先使用替代拆分变量（如果有）。（替代拆分是使用替代预测变量尽可能匹配原始拆分的拆分。）如果没有指定替代拆分或者所有替代拆分变量都缺失，则使用最大的子节点。
- **Logistic 回归模型。**对于 Logistic 回归模型中的协变量，如果预测变量的均值作为保存的模型的一部分包含在其中，则在评分计算中将使用此均值代替缺失值，评分过程继续。如果预测变量是分类变量（例如，Logistic 回归模型中的因子），或者如果均值不可用，则返回系统缺失值。

使用系统缺失值。对具有缺失值的个案进行评分时返回系统缺失值。

选择评分函数

评分函数是所选模型可用的“得分”类型。例如，目标的预测值、预测值的概率或所选目标值的概率。

图片 15-3
评分向导：选择评分函数



评分函数。 可用的评分函数取决于模型。下列值中的一个或多个将在列表中可用：

- **预测值。** 兴趣目标结果的预测值。这对所有模型可用，没有目标的模型除外。
- **预测值的概率。** 预测值的概率是以比例表示的正确值。这对具有分类目标的大部分模型可用。
- **所选值的概率。** 所选值的概率是以比例表示的正确值。从“值”列中的下拉列表选择一个值。可用值由模型定义。这对具有分类目标的大部分模型可用。
- **置信。** 与分类目标的预测值关联的概率测量。对于二元 Logistic 回归、多项 Logistic 回归和 Naive Bayes 模型，其结果与预测值的概率相同。对于树和 Ruleset 模型，置信可以被解释为预测类别的调整概率，而且始终比预测值的概率小。对于这些模型，置信值比预测值的概率更加可靠。
- **节点编号。** 树模型的预测终端节点编号。
- **标准误。**预测值的标准误。对带刻度目标的线性回归模型、一般线性模型和广义线性模型可用。这只在模型文件中保存了协方差矩阵时可用。
- **累积风险**. 估计累积风险函数。该值指示了在给定预测变量值的前提下，在指定时间或该时间之前观察到事件的概率。
- **最近邻元素。** 最近邻元素的 ID。如果提供的话，该 ID 是个案标签变量的值，否则则为个案编号。只应用于最近邻元素模型。

- **第 K 个最近邻元素。** 第 k 个最近邻元素的 ID。在“值”列中输入一个整数作为 k 值。如果提供的话，该 ID 是个案标签变量的值，否则则为个案编号。只应用于最近邻元素模型。
- **到最近邻元素的距离。** 到最近邻元素的距离。根据不同模型，将使用 Euclidean 或城市街区距离。只应用于最近邻元素模型。
- **到第 k 个最近邻元素的距离。** 到第 k 个最近邻元素的距离。在“值”列中输入一个整数作为 k 值。根据不同模型，将使用 Euclidean 或城市街区距离。只应用于最近邻元素模型。

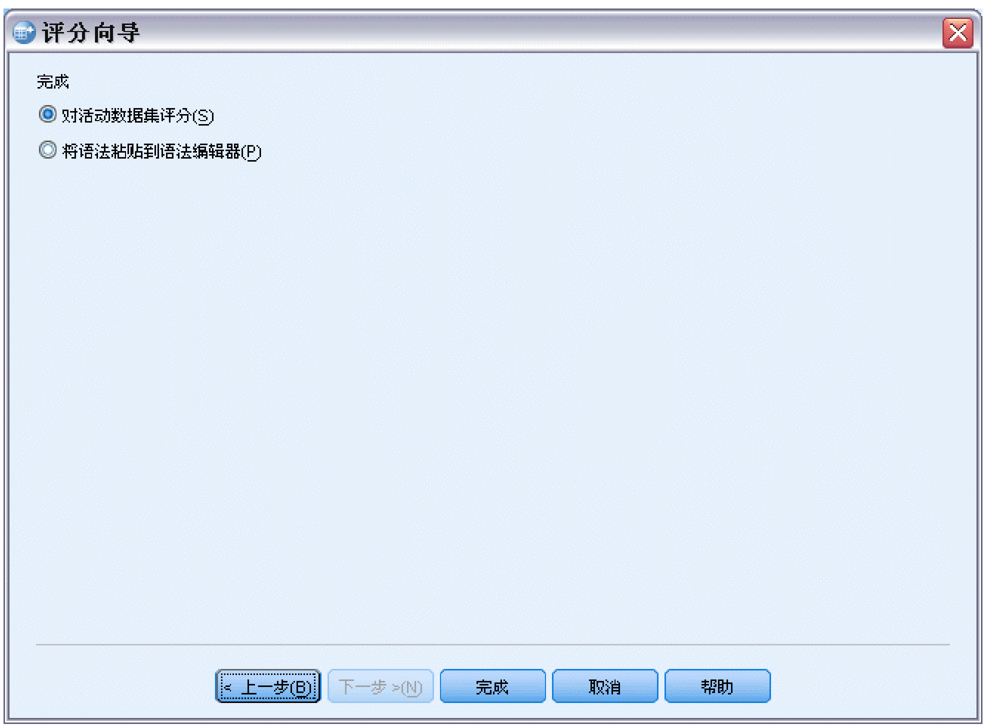
字段名称。 每个选定的评分函数在活动数据集中保存一个新的字段（变量）。您可以使用默认名称或输入新名称。如果活动数据集中已存在具有相同名称的字段，它们将被替换。有关字段命名规则的信息，请参见 变量名。

值。 有关使用“值”设置的函数的说明，请参见评分函数的说明。

对活动数据集进行评分

在向导的最后步骤，您可以对活动数据集进行评分或粘贴所生成的命令语法到语法窗口。然后，您可以修改和/或保存所生成的命令语法。

图片 15-4
评分向导：完成



合并模型和转换 XML 文件

一些预测模型使用以不同方式修改或转换的数据构建。为了以有意义的方式将这些模型应用到其他数据集，还必须在要评分的数据集上执行相同的转换，或将该转换同时反映在模型文件中。在模型文件中包括转换的过程分为两个步骤：

- 保存转换到转换 XML 文件中。这只能通过在命令语法中使用 TMS BEGIN 和 TMS END 来完成。
- 将模型 XML 文件和转换 XML 文件合并到一个新的合并模型 XML 文件中。

要将模型 XML 文件和转换 XML 文件合并到新的合并模型文件：

- 从菜单中选择：

实用程序 > 合并模型 XML

图片 15-5
“合并模型 XML”对话框



- 选择模型 XML 文件。
- 选择转换 XML 文件。
- 输入新的合并模型 XML 文件的路径和名称，或使用浏览选择位置和名称。

注意： 您无法合并模型 ZIP 存档与转换 XML 文件。您只能合并模型 XML 文件和转换 XML 文件。

章

16

实用程序

本章描述可在“实用程序”菜单上找到的功能，以及重新排列目标变量列表的能力。

- 有关评分向导的信息，请参见使用预测模型对数据评分。
- 有关合并模型和转换 XML 文件的信息，请参见合并模型和转换 XML 文件。

变量信息

“变量”对话框为当前选定的变量显示变量定义信息，包括：

- 变量标签
- 数据格式
- 用户缺失值
- 值标签
- 测量级别

图片 16-1
“变量”对话框



可视。变量列表中的“可视”列指示变量当前在数据编辑器或对话框变量列表中是否可见。其可见性由变量集控制。

转到。转至数据编辑器窗口中选定的变量。

粘贴。将选定的变量粘贴到指定语法窗口中光标所在的位置。

要修改变量定义，可使用数据编辑器的变量视图。

获取变量信息

▶ 从菜单中选择：
实用程序 > 变量...

▶ 选择想要显示其变量定义信息的变量。

数据文件注释

您可以随数据文件包含描述性注释。对于 IBM® SPSS® Statistics 数据文件，这些注释随数据文件一起保存。

添加、修改、删除或显示数据文件注释

▶ 从菜单中选择：
实用程序 > 数据文件注释...

▶ 要在“查看器”中显示注释，请选择**在输出中显示注释**。

注释可以是任何长度，但每行最多 80 个字节（在单字节语言中通常为 80 个字符），满 80 个字符将自动换行。注释的显示字体与文本输出的字体相同，精确地反映它们将在“查看器”中如何显示。

只要添加或修改注释，日期戳记（括号中的当前日期）就会自动附加到注释列表的末尾。如果您修改现有注释，或在现有注释之间插入新注释，这可能混淆与注释相关联的日期。

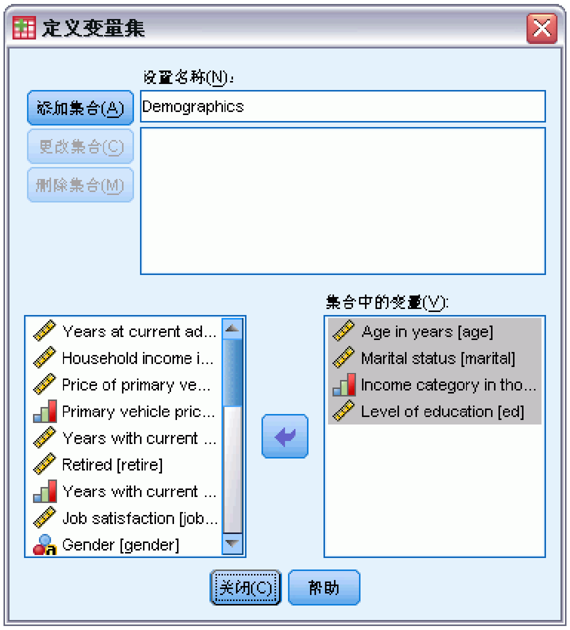
变量集

您可以通过定义和使用变量集，限制数据编辑器和对话框变量列表中显示的变量。对于含有大量变量的数据文件，这一点尤其有用。小型变量集更容易找到和选择变量供您分析。

定义变量集

“定义变量集”可创建要显示在数据编辑器和对话框变量列表中的变量子集。定义的变量集保存在 IBM® SPSS® Statistics 数据文件中。

图片 16-2
“定义变量集”对话框



集合名称。集合名称最多可以包含 64 个字节。可以使用任意字符，包括空白。

集合中的变量。集合中可包含数值变量和字符串变量的任何组合。变量在集合中的顺序对变量在数据编辑器或对话框变量列表中显示的顺序没有任何影响。一个变量可以属于多个集合。

定义变量集

▶ 从菜单中选择：

实用程序 > 定义变量集...

▶ 选择想要在集合中包含的变量。

▶ 为集合输入一个名称（不超过 64 个字节）。

▶ 单击添加集合。

使用变量集合显示和隐藏变量

“使用变量集”将数据编辑器和对话框变量列表中显示的变量限制为所选（选中）集合中的变量。

图片 16-3
“使用变量集”对话框



- 数据编辑器和对话框变量列表中显示的变量集合是所有选择的集合的并集。
- 一个变量可以包含在多个选定集合中。
- 变量在选定集合中的顺序以及选定集合的顺序对变量在数据编辑器或对话框变量列表中显示的顺序没有任何影响。
- 尽管定义的变量集保存在 IBM® SPSS® Statistics 数据文件中，但当前选择的集的列表在您每次打开数据文件时都重置为默认内置集。

可用变量集列表包括为活动数据集定义的所有变量集，加上两个内置集：

- **ALLVARIABLES**。此集合包含数据文件中的所有变量，包括在会话期间新建的变量。
- **NEWVARIABLES**。此集合仅包含在会话期间新建的变量。

  注意：即使在创建新变量后保存数据文件，这些新变量仍然包括在 NEWVARIABLES 集合中，直到您关闭并重新打开数据文件。

必须至少选择一个变量集。如果选择 ALLVARIABLES，则任何其他选择的集将没有任何视觉效果，因为该集包含所有变量。

选择要显示的变量集

▶ 从菜单中选择：
实用程序 > 使用变量集...

▶ 选择定义的变量集，其中包含您想要显示在数据编辑器和对话框变量列表中的变量。

显示所有变量

▶ 从菜单中选择：
实用程序 > 显示所有变量

重新排序目标变量列表

变量在对话框目标列表中出现的顺序就是从源列表中选择它们的顺序。如果想要更改变量在目标列表上的顺序，但又不想取消选择所有变量并以新的顺序重新选择它们，您可以使用 Ctrl 键（Macintosh：Command 键）和向上与向下箭头按键将变量在目标列表中上下移动。如果多个变量连续（组在一起），您可以同时移动它们。不能移动不连续的变量组。

使用扩展束

扩展束能够对定制组件进行打包，例如定制对话框和扩展命令，以方便最终用户安装。例如，如果您为某个扩展命令创建了定制对话框，并打算共享给其他用户，这时您将需要创建包含定制对话框包装 (.spd) 文件和扩展命令关联文件（指定扩展命令语法的 XML 文件，以及采用 Python 或 R 编写的实现代码文件）的扩展束。然后，您就可以将扩展束共享给其他用户。

创建扩展束

如何创建扩展束

▶ 从菜单中选择：
实用程序 > 扩展束 > 创建扩展束...

▶ 在“必需”选项卡上为所有字段输入值。

▶ 在“可选”选项卡上，为您的扩展束需要的字段输入值。

▶ 为扩展束指定目标文件。

▶ 单击**保存**将扩展束保存到指定的位置。这将关闭“创建扩展束”对话框。

扩展束的必需字段

名称。与扩展束关联的唯一名称。它可以包含最多三个单词，且不区分大小写。仅限使用 7 位 ASCII 字符。为降低名称冲突可能性，您可能需要使用多个单词的名称，其中首个单词包含您组织的标识（例如 URL）。

文件。单击**添加**添加与扩展束关联的文件。一个扩展束必须至少包含扩展命令的一个定制对话框规范 (.spd) 文件或一个 XML 规范文件。如果包含有 XML 规范文件，则扩展束还必须至少包含一个 Python 或 R 代码文件，具体来说即 .py、pyc、.pyo 或 .R 类型的文件。

- 在扩展束中包含的扩展命令（采用 Python 或 R 实现）的转换文件将从“可选”选项卡上的“转换目录文件夹”字段中添加。
- 您可以在扩展束中添加自述文件。将文件名指定为 ReadMe.txt。最终用户将能够从显示扩展束详细信息的对话框中访问该自述文件。您可以包含自述文件的本地化版本，并将其指定为 ReadMe_<语言标识符>.txt，例如法语版本的文件名为 ReadMe_fr.txt。

摘要。关于扩展束的简短描述，用于在单行中显示。

描述。关于扩展束的描述说明，它应比“摘要”字段中的描述更加详细。例如，您可以列出扩展束提供的主要功能。如果扩展束针对来自 R 程序包的函数提供了包装器，则应在此处加以说明。

作者。扩展束的作者。您可能希望包括一个电子邮件地址。

版本。以 x.x.x 形式提供的版本标识，其中每个部分必须为整数，例如 1.0.0，如果未提供，则表示为零。例如，版本标识 3.1 会表示为 3.1.0。

最低 SPSS Statistics 版本。运行扩展束所需的最低 IBM® SPSS® Statistics 版本。

扩展束的可选字段

发布日期。发布日期为可选。无格式规定。

链接。与扩展束关联的一组可选 URL，例如，作者的主页。此字段的格式不受限制，因此请确保使用空格、逗号或其他有效的分隔符来分隔多个 URL。

类别。与扩展束关联的一组可选类别，在创作扩展束以供发布到 Developer Central（http://www.spss.com/devcentral扩展命令、统计量和 R。

所需插件。选中运行与扩展束关联的定制组件所需的任何插件（Python 或 R）的复选框。如果缺少所需插件，会在安装期间向用户发出警告。

所需 R 程序包。输入扩展束需要的来自 CRAN 程序包库的任何 R 程序包名称。名称区分大小写。要添加第一个程序包，在“所需 R 程序包”控件中单击任意位置以突出显示输入字段。保持光标在给定行中，按 Enter 键将创建新行。选中某行并按 Delete 键可删除此行。在安装扩展束后，IBM® SPSS® Statistics 将检查所需 R 程序包是否在最终用户的计算机上存在，并尝试下载和安装任何缺失的程序包。

所需 Python 模块。输入扩展束需要但尚未添加的任何 Python 模块名称。所有此类模块应发布到 Developer Central（http://www.spss.com/devcentralSPSS Statistics 安装目录下的 extensions 目录中，或由 SPSS_EXTENSIONS_PATH 环境变量指定的备用位置。

转换目录文件夹。您可以包含一个存放转换目录的文件夹。这允许您为扩展束所包含扩展命令的实现程序（Python 或 R）提供本地化消息和本地化输出。浏览至包含所需转换目录的 lang 文件夹，并选择此文件夹。有关本地化 Python 和 R 程序输出的详细信息，请参阅帮助系统中 Python 集成插件和 R 集成插件相关主题。

注意：自定义对话框的本地化需要不同的机制。

安装扩展束

安装扩展束：

- ▶ 从菜单中选择：
  实用程序 > 扩展束 > 安装扩展束...
- ▶ 选择所需的扩展束。扩展束的文件类型为 spe。

如果扩展束包含任何扩展命令，您将需要重新启动 IBM® SPSS® Statistics 以使用这些命令。要查看当前已安装的扩展束的详细信息，请选择“实用程序 > 扩展束 > 查看已安装的扩展束”。

您还可以使用命令行实用程序安装扩展束，这样可以一次安装多个扩展束。

在分布模式下安装扩展束

如果您在分布模式下工作，则需要使用命令行实用程序，将扩展束安装在相关的 SPSS Statistics 服务器上。如果扩展束包含自定义对话框，则还必须在本地计算机上安装扩展束（如上所述，从菜单中安装）。如果您不知道扩展束是否包含自定义对话框，则最好在 SPSS Statistics 服务器计算机和本地计算机上同时安装扩展束。有关使用命令行实用程序的详细信息，请参见扩展束的批安装第 287 页码。

扩展束的安装位置

对于 Windows 和 Linux，默认情况下，安装扩展束需要对 IBM® SPSS® Statistics 安装目录具有写入权限（对于 Mac，扩展束安装在常规用户可写入的位置）。如果对所需位置没有写入权限，或打算在其它位置保存扩展束，则可通过定义 SPSS_EXTENSIONS_PATH 环境变量以指定一个或多个备用位置。如果在 SPSS_EXTENSIONS_PATH 中指定了路径，则它优先于默认位置。扩展束将安装在第一个可写入的位置。注意，Mac 用户还可以使用 SPSS_EXTENSIONS_PATH 环境变量。如果要指定多个备用位置，对于 Windows 请使用分号分隔，对于 Linux 和 Mac，则使用冒号分隔。指定位置必须在目标计算机上存在。在设置 SPSS_EXTENSIONS_PATH 后，您将需要重启 SPSS Statistics 以使更改生效。

如果扩展束包含定制对话框且您不具有 SPSS Statistics 安装目录的写入权限（对于 Windows 和 Linux），则您将需要设置 SPSS_CDIALOGS_PATH 环境变量。SPSS_CDIALOGS_PATH 的规则也适用于 SPSS_EXTENSIONS_PATH。如果您不知道当前安装的扩展束是否包含定制对话框，并且您对默认位置没有写入权限，那么最好在 SPSS_CDIALOGS_PATH 中指定一个备用位置。

要在 Windows 上创建 SPSS_EXTENSIONS_PATH 或 SPSS_CDIALOGS_PATH 环境变量，从“控制面板”中：

Windows XP

- ▶ 选择“系统”。
- ▶ 选择“高级”选项卡并单击**环境变量**。
- ▶ 在“用户变量”部分，单击**新建**，在“变量名”字段中输入环境变量的名称，例如 SPSS_EXTENSIONS_PATH，并在“变量值”字段中输入所需路径。

Windows Vista 或 Windows 7

- ▶ 选择“用户帐户”。
- ▶ 选择“更改我的环境变量”。

▶ 单击**新建**，在“变量名”字段中输入环境变量名称—如 SPSS_EXTENSIONS_PATH —并在“变量值”字段中输入所需路径。

要查看扩展束和定制对话框的当前位置（扩展束与扩展命令的位置相同），运行以下命令语法：SHOW EXTPATHS。

所需 R 程序包

扩展束安装程序会尝试下载和安装扩展束需要但在您的计算机上未找到的 R 程序包。如果您的计算机无法访问 Internet，则需要向其他拥有所需程序包的人员索取。您可以从 http://www.r-project.org/ 下载程序包，然后在 R 中安装。有关详细信息，请参见随 R 分发的 R 安装与管理指南。如果安装程序包失败，将显示所需程序包列表。在扩展束安装后，您还可以从“扩展束详细信息”对话框中查看此列表。

注意：对于 UNIX（包括 Linux）用户，程序包以源代码形式下载，需要进行编译。因此必须在您的计算机上安装相应的工具。请参见 R 安装与管理指南获得详细信息。具体而言，Debian 用户应当通过 apt-get install r-base-dev 安装 r-base-dev 程序包。

权限

默认情况下，所需的 R 程序包将被安装到 R 安装位置下的 library 文件夹中，例如，Windows 上的 C:\Program Files\R\R-2.10.1\library。如果对此位置没有写入权限，或打算在其他位置保存为扩展束安装的 R 程序包，则可通过定义 SPSS_RPACKAGES_PATH 环境变量以指定一个或多个备用位置。如果在 SPSS_RPACKAGES_PATH 中指定了路径，则它优先于默认位置。R 程序包将安装在第一个可写入的位置。如果要指定多个备用位置，对于 Windows 请使用分号分隔，对于 Linux 和 Mac，则使用冒号分隔。指定位置必须在目标计算机上存在。在设置 SPSS_RPACKAGES_PATH 后，您将需要重启 IBM® SPSS® Statistics 以使更改生效。有关如何在 Windows 上设置环境变量的信息，请参见扩展束的安装位置第 286 页码。

扩展束的批安装

您可以使用位于 IBM® SPSS® Statistics 安装目录下的批处理实用程序 installextbundles.bat（在 Mac 和 UNIX 系统中为 installextbundles.sh），一次安装多个扩展束。对于 Windows 和 Mac，该实用程序位于安装目录的根位置。对于 Linux 和用于 UNIX 版本的 SPSS Statistics 服务器，该实用程序位 安装目录的 bin 子目录。该实用程序设计为从命令提示符运行，且必须从其安装位置运行。命令行格式为：

```
installextbundles [-statssrv] [-download no|yes ] -source <folder> | <filename>...
```

-statssrv。指定您在 SPSS Statistics 服务器上运行此实用程序。您还需要在连接到服务器的客户端计算机上安装相同的扩展束。

-download no|yes。指定此实用程序是否有权访问 Internet，以便下载指定扩展束所需的任何 R 程序包。默认值为 no。如果保留默认值，或无法访问 Internet，您将需要手动安装任何所需的 R 程序包。

- source <folder> | <filename>... 指定要安装的扩展束。您可以指定包含扩展束的文件夹路径，或指定扩展束文件名的列表。如果指定文件夹，则会安装在此文件夹中找到的所有扩展束（文件类型 .spe）。多个文件名之间以一个或多个空格分开。如果路径包含空格，则需要用双引号引起。

注意：当在用于 UNIX 版本的 SPSS Statistics 服务器上运行 installextbundles.sh 时，Bash shell 必须存在于服务器计算机上。

查看已安装的扩展束

查看您计算机上已安装扩展束的详细信息：

▶ 从菜单中选择：
实用程序 > 扩展束 > 查看已安装的扩展束...

▶ 针对所需扩展束，单击摘要列中突出显示的文本。

“扩展束详细信息”对话框中显示扩展束作者提供的信息。除必需信息（例如“摘要”、“描述”和“版本”）外，作者还可提供指向相关位置（例如作者的主页）的 URL。扩展束可以包含或不包含自述文件。

组件。“组件”组列出扩展束包含的定制对话框（如果有）和任何扩展命令的名称。扩展束包含的扩展命令可以从语法编辑器中运行，其运行方式与内置 IBM® SPSS® Statistics 命令相同。在语法编辑器中，运行 **CommandName /HELP** 可以获得扩展命令的帮助。如果定制对话框包含帮助按钮，则也可访问其帮助。

注意：安装包含定制对话框的扩展束需要重新启动 SPSS Statistics，以便在“组件”组中查看对话框的输入。

相关性。“相关性”组列出运行扩展束中包含的组件所需的插件。

- **Python 和 R 的集成插件。** 扩展束的组件可能需要 Python 的集成插件和/或 R 的集成插件，二者均为免费提供。
- **R 程序包。**列出扩展束所需的任何 R 程序包。在扩展束安装期间，安装程序尝试下载并在您的计算机上安装所需的程序包。如果此过程失败，将提醒您手动安装程序包.
- **Python 模块。**列出扩展束所需的任何 Python 模块。所有此类模块可从 Developer Central (http://www.spss.com/devcentralSPSS Statistics 安装目录下的 extensions 目录。对于 Mac，这些模块应当复制到 /Library/Application Support/IBM/SPSS/Statistics/19/extensions 目录。如果您已使用 SPSS_EXTENSIONS_PATH 环境变量为扩展束指定了备用位置，则应将 Python 模块复制到某个备用位置。或者，您还可以将模块复制到 Python 搜索路径中的某个位置，例如 Python site-packages 目录。

章
17

选项

选项控制各种设置，包括：

- 会话日志，记录每次会话中运行的所有命令
- 显示对话框源列表中的变量的顺序
- 在新输出结果中显示的项和隐藏的项
- 新枢轴表的表格外观
- 自定义货币格式

更改选项设置

▶ 从菜单中选择：
编辑 > **选项...**

▶ 单击要更改设置的选项卡。

▶ 更改设置。

▶ 单击**确定**或**应用**。

一般选项

图片 17-1
“选项”对话框：“常规”选项卡



变量列表

这些设置控制对话框列表中变量的显示。可以显示变量名或变量标签。名称或标签可按字母顺序或文件顺序显示，或按测量级别进行分组。显示顺序仅影响源变量列表。目标变量列表总是反映变量选择的顺序。

角色

某些对话框支持基于定义的角色预先选择分析变量的功能。

- **使用预定义角色。** 默认情况下，基于定义的角色预先选择变量。
- **使用自定义分配。** 默认情况下，不要使用角色预先选择变量。

在支持此功能的对话框中，您还可以在预定义角色和自定义分配之间切换。此处的设置仅控制每个数据集的初始默认有效行为。

Windows

观感。控制窗口和对话框的基本外观。如果您在更改外观后发现任何显示问题，请尝试关闭并重新启动应用程序。

在启动时打开语法窗口。语法窗口是用于输入、编辑和运行命令的文本文件窗口。如果您经常使用命令语法，请选择该选项，在每次会话开始时自动打开语法窗口。对于喜欢用命令语法而不喜欢用对话框的熟练用户来说，此选项很有用。

一次只能打开一个数据集。 每次使用菜单和对话框打开其他数据源时，请关闭当前打开的数据源。默认情况下，每次使用菜单和对话框打开新数据源时，该数据源会在新的“数据编辑器”窗口中打开，并且在其他“数据编辑器”窗口中打开的任何其他数据源都会保持打开状态，并在会话期间可用，直到明确关闭。

选择此选项会立即生效，但不会关闭在更改设置时所打开的任何数据集。该设置对使用依赖于 DATASET 命令来控制多个数据集的命令语法所打开的数据源没有任何影响。

数据文件和语法文件的字符编码

该选项控制确定读写数据文件和语法文件的编码方式的默认行为。您只能在未打开数据源时更改此设置，并且该设置对后续会话继续有效，直到明确更改。

- **Locale 的写入系统。**使用当前区域设置确定读写文件的编码方式。这也称为**代码页模式**。
- **Unicode（通用字符集）。**使用 Unicode 编码 (UTF-8) 来读写文件。这也称为 Unicode 模式。

有关 Unicode 模式和 Unicode 文件的一些重要隐含信息：

- 以 Unicode 编码保存的 IBM® SPSS® Statistics 数据文件 和语法文件 不能在 16.0 以前版本的 SPSS Statistics 上使用。对于语法文件，您可在保存文件时指定编码方式。 对于数据文件，如果要使用较早的版本读取文件，则应该以代码页模式打开数据文件，然后重新保存。
- 以 Unicode 模式读取代码页数据文件时，所有字符串变量的限定宽度扩大为原来的三倍。要将每个字符串变量的宽度自动设置为该变量的最长观察值，请选择“打开数据”对话框中的**根据观察值来最小化字符串宽度**。

输出

表格中较小的数值没有科学计数法。对于输出中较小的小数值不显示科学计数法。非常小的小数值将显示为 0（或者 0.000）。

将区域的数字分组格式应用到数值。 将当前的区域’数字分组格式应用到枢轴表和图表中的数值以及数据编辑器中的数值。例如，在启用了该设置的法语区域设置中，值 34419,57 将显示为 34 419,57。

分组格式不适用于树、模型查看器项目、DOT 或 COMMA 格式的数值以及 DOLLAR 或自定义货币格式的数值。但它适用于以 DTIME 格式数值来显示日期值—例如，以 ddd hh:mm 的格式显示 ddd 的值。

测量系统。用于指定属性（如枢轴表单元格边距、单元格宽度和打印时表之间的空白）的测量系统（点、英寸或厘米）。

语言。控制输出中使用的语言。不适用于简单文本输出。可用语言的列表取决于当前安装的语言文件。（注意：这不会影响用户界面语言。）根据所选的语言，您可能还需要使用 Unicode 字符编码以正确显示字符。

注意：输出时依赖于语言特定的文本字符串的自定义脚本在更改输出语言后可能无法正常运行。

提示。控制程序通知您已经完成运行过程或者可以在查看器中获得结果的方式。

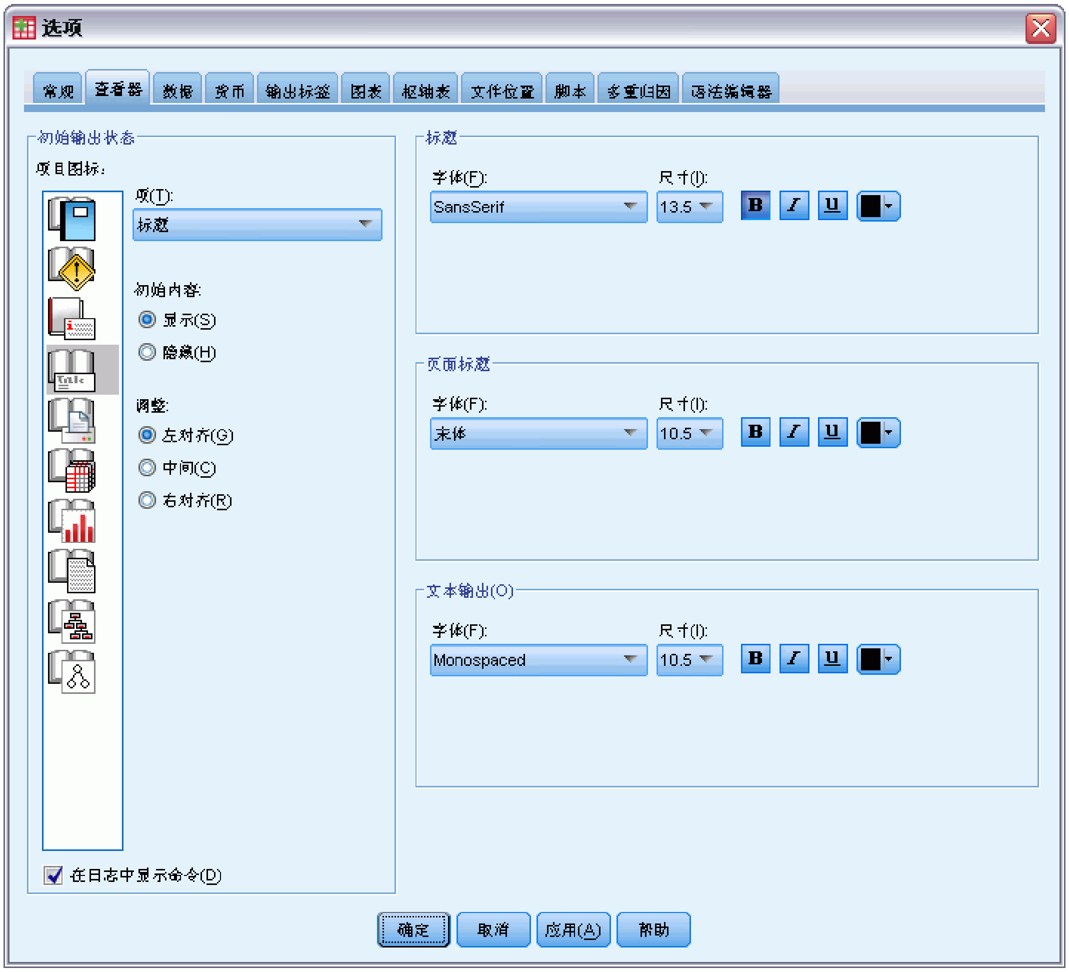
用户界面

它控制菜单、对话框和其他用户界面功能中使用的语言。（注意：这不会影响输出语言。）

查看器选项

查看器输出显示选项仅影响更改设置后生成的新输出。已经显示在查看器中的输出不受这些设置更改的影响。

图片 17-2
“选项”对话框：“查看器”选项卡



初始输出状态。控制每次运行过程时自动显示或隐藏的项，以及各项的初始对齐方式。可以控制以下各项的显示：日志、警告、注释、标题、枢轴表、图表、树形图和文本输出。也可以打开或关闭日志中命令的显示。可以从日志复制命令语法，并将其保存在语法文件中。

注意：在查看器中所有输出项都按照左对齐方式显示。只有打印输出的对齐方式受对齐方式设置的影响。居中和右对齐项用项左上方的一个小符号表示。

标题。控制新输出标题的字体样式、大小和颜色。

页面标题。控制新页面标题以及由 TITLE 和 SUBTITLE 命令语法生成的或由“插入”菜单上的**新页面标题**创建的页面标题的字体样式、大小和颜色。

用于文本输出的**文本输出**字体。文本输出设计成与等宽（固定间距）字体一起使用。如果选择比例字体，表格输出将不会正确对齐。

数据选项

图片 17-3
“选项”对话框：“数据”选项卡



转换与合并选项。程序每次执行命令时，都要读取数据文件。有些数据转换（如“计算”和“重新编码”）和文件转换（如“添加变量”和“添加个案”）不要求单独传送数据，这些命令的执行可以延迟到程序读取数据以执行另一个命令时为止，如统计或绘图过程。

- 对于读取数据消耗一定时间的大型数据文件，可以选择**用前评价**来延迟执行并节省处理时间。选择此选项后，使用类似“计算变量”的对话框进行的转换的结果将不会立即显示在数据编辑器中；转换创建的新变量显示时将不带任何数据值；存在挂起的转换时，不能更改数据编辑器中的数据值。任何读取数据的命令（如统计或绘图过程）都将执行挂起的转换并更新数据编辑器中显示的数据。或者，可以使用“转换”菜单上的**运行挂起的转换**。
- 如果使用默认设置**立即评价**，则从对话框粘贴命令语法时，每次转换命令后都将粘贴 EXECUTE 命令。

显示新数值变量的格式。控制新数值变量的默认显示宽度和小数位数。新字符串变量没有默认的显示格式。如果值对于指定的显示格式来说太大了，则第一个小数位进行四舍五入，然后值就转换为科学计数法。显示格式不影响内部数据值。例如，数值 123456.78 在显示时可能四舍五入为 123457，但是在任何计算中都会使用原始的未经过四舍五入的值。

设置用两位数字表示年号的年限全距。为用两位数年份（如 10/28/86，29-OCT-87）输入和/或显示的日期格式变量定义年份范围。自动范围设置是基于当年的，从 69 年前开始，到 30 年后结束（加上当年，整个范围为 100 年）。对于自定义范围，结束年份是根据您输入的开始年份的值自动确定的。

随机数字生成器。有两种不同的随机数字生成器可供使用：

- **兼容版本 12.** 版本 12 和早期版本中所使用的随机数字生成器。如果您要重新生成在早期版本中根据指定的种子值生成的随机结果，可使用此随机数字生成器。
- **Mersenne 扭曲器.** 较新的随机数生成器，可实现更可靠的模拟。如果从版本 12 或早期版本再现随机结果不是问题，那么可使用此随机数生成器。

读取外部数据。 对于从外部文件格式读取的数据和较旧版本的 IBM® SPSS® Statistics 数据文件（8.0 以前的版本），您可以指定数值变量的数据值的最小数量，用于将变量分为连续（刻度）变量或名义变量。唯一值的个数少于指定数量的变量划分为名义变量。

当决定应用连续（刻度）或名义测量级别时，会在应用最小数据值规则数之前评估许多其他条件。将以条件在下表格中的排列顺序对其进行评估。将应用与数据匹配的第一个条件的测量级别。

条件	测量级别
格式为美元或定制货币	连续
格式为日期或时间（不包括月份和星期）	连续
变量的所有值均缺失	名义
变量包含至少一个非整数值	连续
变量包含至少一个负值	连续
变量不包含少于 10,000 的有效值	连续
变量具有 N 个或更多唯一有效值*	连续
变量不包含少于 10 的有效值	连续
变量具有少于 N 个唯一有效值*	名义

* N 是用户指定的临界值。默认值为 24。

数值的四舍五入和取整。 对于 RND 和 TRUNC 函数，此设置控制用来将接近四舍五入边界的值四舍五入的默认阈值。在安装时，此设置被指定为一个位数，设为 6，这对于大部分应用程序都足够了。位数设为 0 会产生与版本 10 相同的结果。将位数设为 10 会产生与版本 11 和 12 相同的结果。

- 对于 RND 函数，此设置指定最不显著的位数，要四舍五入的值可能未达到四舍五入的阈值，但是仍然被四舍五入。例如，在将一个 1.0 到 2.0 之间的值四舍五入为最接近的整数时，此设置会指定该值距离 1.5 还差多少（四舍五入为 2.0 的阈值），并仍然将该值四舍五入为 2.0。
- 对于 TRUNC 函数，此设置指定最不显著的位数，要取整的值可能未达到最接近的四舍五入边界，在取整之前要被四舍五入。例如，在将一个 1.0 到 2.0 之间的值取整为最接近的整数时，此设置指定该值距离 2.0 还差多少，并将该值四舍五入为 2.0。

自定义变量视图。控制变量视图中属性的默认显示和顺序。

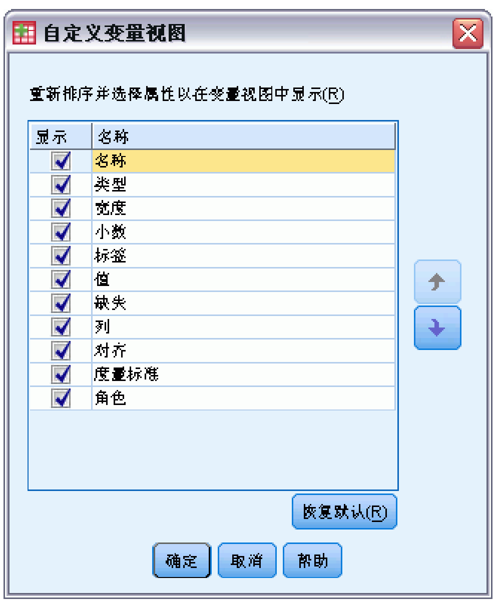
更改字典。控制用于检查变量视图中项目拼写的字典的语言版本。

更改默认变量视图

默认情况下，您可以使用“自定义变量视图”来控制“变量视图”中显示的属性（例如，名称、类型、标签）及其显示顺序。

单击**自定义变量视图**。

图片 17-4
自定义变量视图（默认）



▶ 选择（选中）要显示的变量属性。

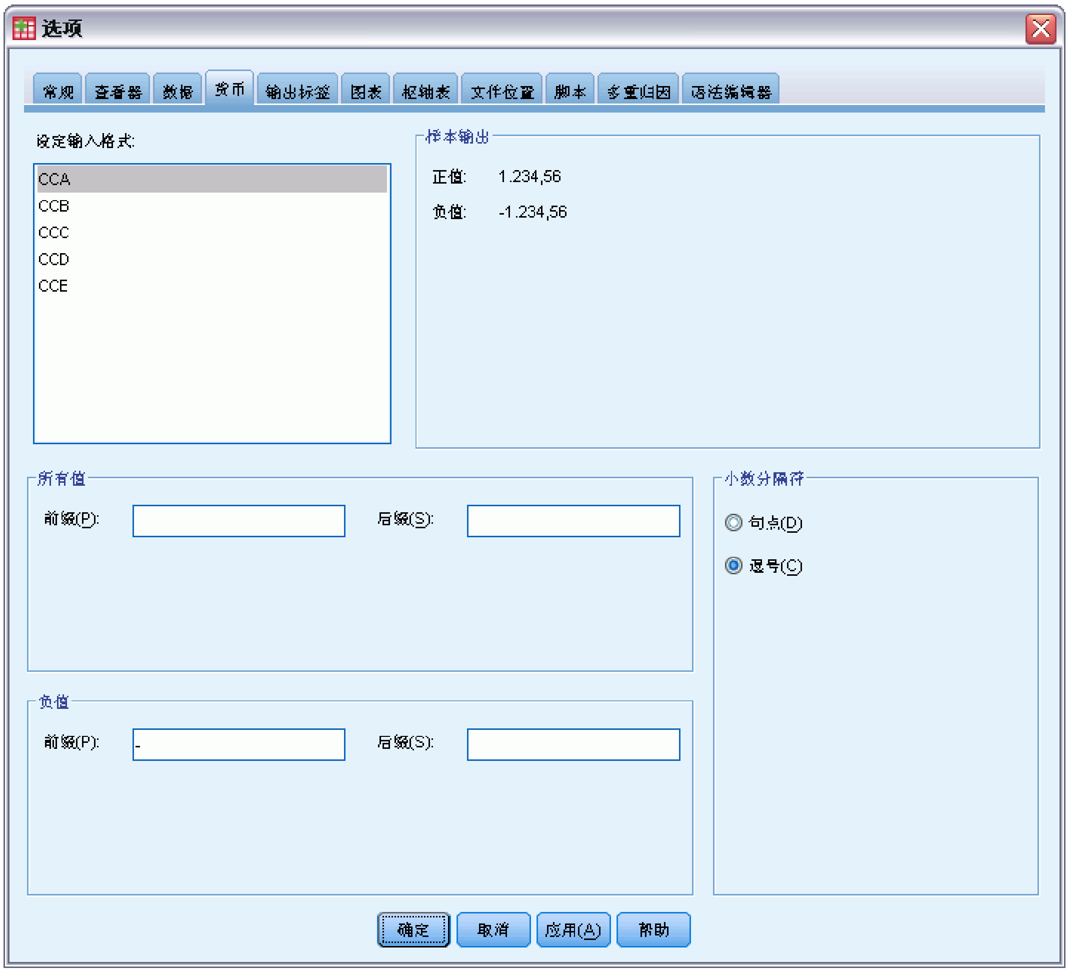
▶ 使用向上和向下箭头按钮更改属性的显示顺序。

货币选项

最多可以创建五种自定义货币显示格式，格式可以包括特殊的前缀和后缀字符以及对负值的特殊处理方式。

五种自定义货币格式名称为 CCA、CCB、CCC、CCD 和 CCE。您不能更改格式名或添加新的格式名。要修改自定义货币格式，请从源列表中选择格式名，然后进行所需的更改。

图片 17-5
“选项”对话框：“货币”选项卡



为自定义货币格式定义的前缀、后缀和小数指示符仅用于显示目的。您不能用自定义货币字符在数据编辑器中输入值。

创建自定义货币格式

▶ 单击**货币**选项卡。

▶ 从列表（CCA、CCB、CCC、CCD 和 CCE）中选择一种货币格式。

▶ 输入前缀、后缀和小数指示符的值。

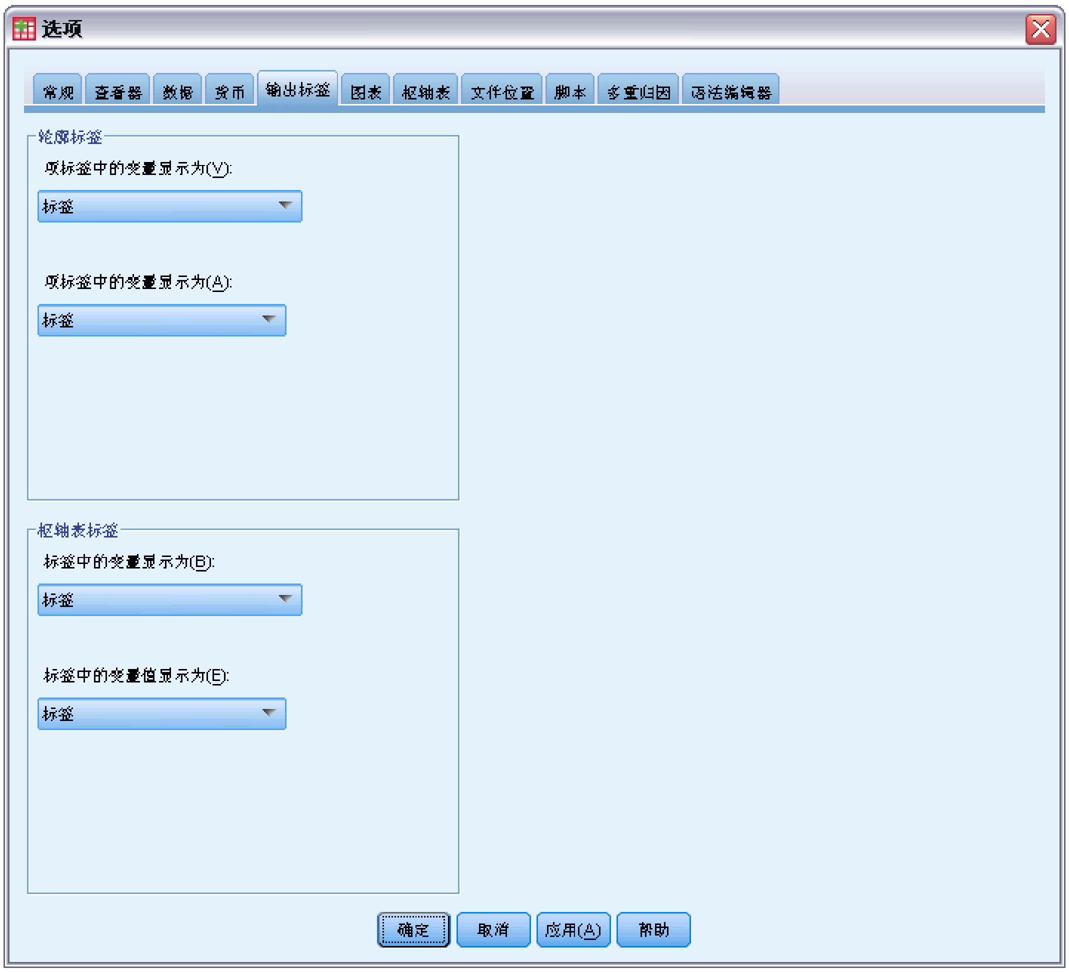
▶ 单击**确定**或**应用**。

输出标签选项

“输出标签”选项控制变量和数据值信息在概要和枢轴表中的显示。可以显示变量名、定义的变量标签和实际数据值、定义的值标签或各项的组合。

描述性变量标签和值标签（数据编辑器、标签和值列中的变量视图）通常可使结果更容易看懂。但是，长标签在一些表中可能很难处理。

图片 17-6
“选项”对话框：“输出标签”选项卡



输出标签选项仅影响更改设置后生成的新输出。已经显示在查看器中的输出不受这些设置更改的影响。这些设置仅影响枢轴表输出。文本输出不受这些设置的影响。

图表选项

图片 17-7
“选项”对话框：“图表”选项卡



图表模板。新的图表可以使用此处选择的设置或图表模板文件中的设置。单击**浏览**，选择一个图表模板文件。要创建图表模板文件，请用所需的属性创建一个图表并将其保存为模板（从“文件”菜单选择**保存图表模板**）。

图表宽高比。新图表的外框的宽和高的比率。可以将宽高比指定为 0.1 到 10.0。小于 1 的值使图表的高度比宽度大。大于 1 的值使图表的宽度比高度大。等于 1 的值生成的是正方形图表。图表创建后，其宽高比就无法更改。

当前设置。可用设置包括：

- **字体**。用于新图表中所有文本的字体。
- **样式循环设置**。新图表的颜色和图案的初始分配。**仅在颜色之间循环**仅使用颜色区分图表元素，而不使用图案。**仅在图案之间循环**仅使用线条样式、标记符号或填充图案来区分图表元素，而不使用颜色。

■ **框架**。控制新图表上的内框和外框的显示。

■ **网格线**。控制新图表上的刻度轴网格线和类别轴网格线的显示。

样式循环。自定义新图表的颜色、线条样式、标记符号和填充图案。可以更改在创建新图表时使用的颜色和图案的顺序。

数据元素颜色

指定新图表中的数据元素（如条和标记）应该使用的颜色的顺序。在“图表选项”主对话框的“样式循环的优先选择”组中选择了包括颜色的选项后，就会使用颜色。

例如，如果您创建了有两个组的复式条形图，并且选择了“图表选项”主对话框中的**在颜色之间循环，然后在图案之间循环**，则“分组图表”列表中的最前面两种颜色将用作新图表上的条的颜色。

更改颜色使用的顺序

▶ 选择**简单图表**，然后选择一种颜色，用于没有类别的图表。

▶ 选择**分组图表**，更改有类别的图表的颜色循环。要更改类别的颜色，请选择一个类别，然后从调色板中为该类别选择一种颜色。

根据需要，您可以：

■ 在选定的类别上面插入一个新的类别。

■ 移动选定的类别。

■ 移去选定的类别。

■ 将顺序重置为默认顺序。

■ 选择一种颜色并并单击**编辑**来编辑该颜色。

数据元素线

指定新图表中的线数据元素应使用的样式的顺序。如果图表包括线数据元素，并且在“图表选项”主对话框的“样式循环的优先选择”组中选择了包括图案的选项，就会使用线样式。

例如，如果您创建了有两个组的线图，并且选择了“图表选项”主对话框中的**仅在图案之间循环**，则“分组图表”列表中的最前面两种样式将用作新图表上的线的图案。

更改线样式使用的顺序

▶ 选择**简单图表**，然后选择一种线样式，用于没有类别的线图。

▶ 选择**分组图表**，更改有类别的线图的图案循环。要更改类别的线样式，请选择一个类别，然后从调色板中为该类别选择一种线样式。

根据需要，您可以：

■ 在选定的类别上面插入一个新的类别。

■ 移动选定的类别。

- 移去选定的类别。
- 将顺序重置为默认顺序。

数据元素标记

指定新图表中的标记数据元素应使用的符号的顺序。如果图表包括标记数据元素，并且在“图表选项”主对话框的“样式循环的优先选择”组中选择了包括图案的选项，就会使用标记样式。

例如，如果您创建了有两个组的散点图，并且选择了“图表选项”主对话框中的**仅在图案之间循环**，则“分组图表”列表中的最前面两种符号将用作新图表上的标记。

更改标记样式使用的顺序

▶ 选择**简单图表**，然后选择一种标记符号，用于没有类别的图表。

▶ 选择**分组图表**，更改有类别的图表的图案循环。要更改类别的标记符号，请选择一个类别，然后从调色板中为该类别选择一种符号。

根据需要，您可以：

- 在选定的类别上面插入一个新的类别。
- 移动选定的类别。
- 移去选定的类别。
- 将顺序重置为默认顺序。

数据元素填充

指定新图表中的条和区域数据元素应使用的填充样式的顺序。如果图表包括条或区域数据元素 并且在“图表选项”主对话框的“样式循环的优先选择”组中选择了包括图案的选项，就会使用填充样式。

例如，如果您创建了有两个组的复式条形图，并且选择了“图表选项”主对话框中的**仅在图案之间循环**，则“分组图表”列表中的最前面两种样式将用作新图表上的条填充图案。

更改填充样式使用的顺序

▶ 选择**简单图表**，然后选择一种填充图案，用于没有类别的图表。

▶ 选择**分组图表**，更改有类别的图表的图案循环。要更改类别的填充图案，请选择一个类别，然后从调色板中为该类别选择一种填充图案。

根据需要，您可以：

- 在选定的类别上面插入一个新的类别。
- 移动选定的类别。
- 移去选定的类别。
- 将顺序重置为默认顺序。

枢轴表选项

“枢轴表”选项可设置与枢轴表显示有关的各种选项。

图片 17-8
“选项”对话框：“枢轴表”选项卡



表格外观。从文件列表中选择一个表格外观，并单击**确定**或**应用**。可以使用随 IBM® SPSS® Statistics 一起提供的表格外观，或者在枢轴表编辑器（从“格式”菜单选择**表格外观**）中创建自己的表格外观。

- **浏览**。使您可以从另一个目录中选择表格外观。
- **设置表格外观目录**。使您可以更改默认的表格外观目录。使用**浏览**导航到您要使用的目录，选择该目录中的“表格外观”，然后选择**设置表格外观目录**。

注意：以 SPSS Statistics 的早期版本创建的表格外观不能用于 16.0 或更高版本。

列宽。控制枢轴表中列宽的自动调整。

- **调整标签和数据，除非常大的表以外。** 对于不超过 10,000 个单元格的表，将列宽调整为列标签宽度或最大的数据值宽度中的较大者。对于超过 10,000 个单元格的表，将列宽调整为列标签的宽度。
- **仅调整标签。**将列宽调整为列标签的宽度。这会生成结构更紧凑的表，但宽度超过标签的数据值可能会被截去。
- **调整标签和数据。** 将列宽度调整为：列标签或最大数据值中较大的宽度。这会生成较宽的表，但可以保证显示所有的值。

  注意： 如果您选择更加快速地呈现表格，则将始终调整标签和数据的列宽。

显示行块。 这些设置控制查看器中大型枢轴表的显示。这些设置对打印大型枢轴表或导出输出到外部格式没有影响。(注意：如果您选择快速呈现表格选项，该选项将不可用。)

- **将表格显示为行块。** 表格在查看器中按部分显示，导航控件允许您查看表格的不同部分。
- **待显示行。** 设置要在每个部分中显示的行数。值必须为 10 和 1000 之间的整数。
- **最大单元格数。** 设置要在每个部分中显示的最大单元格数。值必须为 1000 和 100000 之间的整数。
- **短行/孤立行容差。** 控制表格最内层行维度的最大行数，以便在表格的显示视图之间拆分。例如，如果最内层行维度的每个组中有 6 个类别，则指定值 6 可以防止任何组在显示视图之间拆分。值必须为整数，且不能大于“显示的行数”。

有多个因素会影响实际显示的行数，包括：

- 如果在达到“显示的行数”之前先达到了“最大单元格数”，那么将在此时拆分表格。例如，如果“显示的行数”为 1000，而“最大单元格数”为 10000，则具有 20 列的表将显示在 500 个行块中。
- 短行/孤立行容差设置能使实际显示的行数多于或少于“显示的行数”的指定值。

表格呈现。 表格可以全功能枢轴表或轻量表的形式呈现。轻量表可以比全功能枢轴表更快呈现，但它们不支持许多透视和编辑功能，包括：

- 无法透视轻量表。您可以显示多层表的不同层，但无法在维度之间移动表元素（例如，交换行和列）或更改行或列的显示位置。
- 无法编辑轻量表。您无法隐藏行或列、更改列宽度、修改数据值、更改单元格或表格属性，也无法应用其他表格外观。
- 轻量表不支持在表格外观中定义的自定义维数边界属性。所有其他表格外观属性会在创建表格时被接受。
- 由于无法编辑轻量表，“查找和替换”对话框中的“替换”功能将不会替换轻量表中的指定值。

要将轻量表转换为全功能枢轴表：

▶ 右键单击“查看器”窗口中的表格或表格编辑窗口。

▶ 从上下文菜单中，选择**启用编辑**

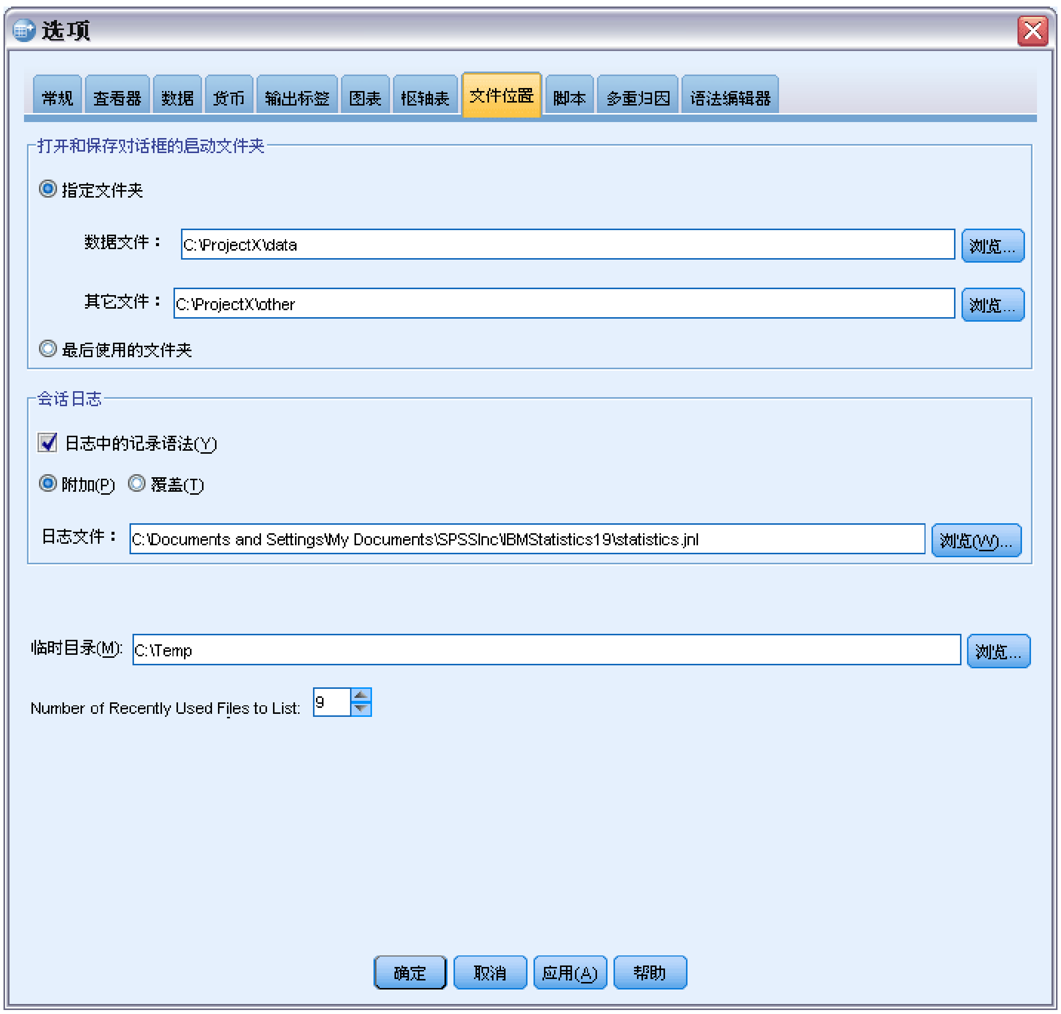
默认编辑模式。控制枢轴表在查看器窗口或独立窗口中的激活。默认情况下，双击枢轴表可激活查看器窗口中的所有除了特大表之外的表格。可以选择在独立窗口中激活枢轴表，或者选择一个大小设置，使小于这个设置的枢轴表在查看器中打开，而大于这个设置的枢轴表在独立的窗口中打开。

以丰富文本格式复制宽表格至剪贴板。当以 Word/RTF 格式粘贴枢轴表时，文档宽度过宽的表格会被换行、按比例缩小以适合文档宽度或不进行更改。

文件位置选项

“文件位置”选项卡上的选项控制应用程序在每个会话开始时打开和保存文件的默认位置、日志文件位置、临时文件夹位置，以及出现在最近使用的文件列表中的文件数量。

图片 17-9
“选项”对话框：“文件位置”选项卡



打开和保存对话框的启动文件夹

- **指定文件夹。**指定的文件夹用作每个会话开头的默认位置。您可为数据和其他（非数据）文件指定不同的默认位置。
- **最后使用的文件夹。**在上一次会话中打开或保存文件的最后一个文件夹，用作下一次会话的默认文件夹。这适用于数据和其他（非数据）文件。

这些设置仅应用于打开和保存文件的对话框—并且“最后使用的文件夹”由用于打开或保存文件的最后一个对话框确定。通过命令语法打开或保存的文件对这些设置无效，且不受这些设置的影响。这些设置仅在本地分析模式下可用。在已连接到远程服务器的分布式分析模式中（需要 IBM® SPSS® Statistics Server），您不能控制启动文件夹位置。

会话日志

您可以使用会话日志自动记录会话中运行的命令。这包括在语法窗口中输入和运行的命令以及由对话框选项生成的命令。可以编辑日志文件，并在其他会话中再次使用命令。可以打开或关闭日志记录，追加或覆盖日志文件，以及选择日志文件的名称和位置。可以从日志文件复制命令语法，并将其保存在语法文件中。

临时文件夹

这将指定在会话过程中创建的临时文件的位置。在分布式模式（随服务器版本一起提供）中，这不会影响临时数据文件的位置。在分布式模式中，临时数据文件的位置由环境变量 SPSSTMPDIR 控制，该变量只能在运行软件的服务器版本的计算机上设置。如果需要更改临时目录的位置，请联系您的系统管理员。

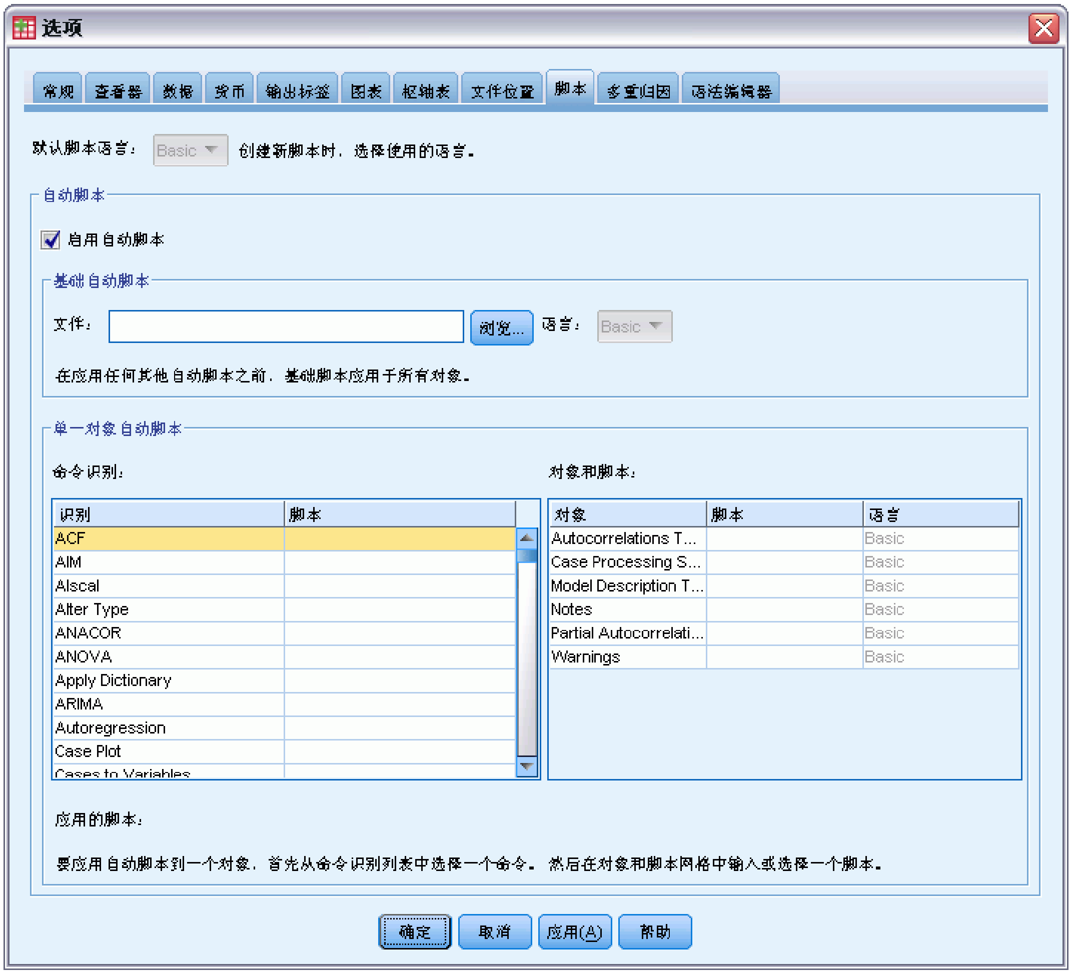
最近使用的文件列表

该选项控制出现在“文件”菜单上的最近使用文件的数量。

脚本选项

使用“脚本”选项卡可以指定默认脚本语言，以及要使用的任何自动脚本。可以使用脚本自动处理许多功能，包括自定义枢轴表。

图片 17-10
“选项”对话框：“脚本”选项卡



注意：Legacy Sax Basic 用户必须手动转换任何自定义自动脚本。随 16.0 以前的版本安装的自动脚本作为一组单独的脚本文件提供，位于 IBM® SPSS® Statistics 安装目录的 Samples 子目录中。默认情况下，输出项与自动脚本无关联。您必须手动将所有自动脚本与所需的输出项相关联，如下所述。有关转换过去遗留自动脚本的信息，请参见与 16.0 之前版本的兼容性第 377 页码。

默认脚本语言。默认脚本语言确定创建新脚本时启动的脚本编辑器。它还指定了其可执行程序将用于运行自动脚本的默认语言。可用脚本语言取决于使用的平台。对于 Windows 平台，可用脚本语言为 Basic（随 Core 系统一起安装）和 Python 编程语言。对于所有其他平台，脚本以 Python 编程语言提供。

要启用使用 Python 编程语言编写脚本的功能，必须安装 Python 和 IBM® SPSS® Statistics - Integration Plug-In for Python。 有关信息，请参阅帮助系统的“Core 系统>常见问题解答”中的“如何获得集成插件”。

启用自动脚本。该复选框允许您启用或禁用自动脚本。默认情况下，已启用自动脚本。

基础自动脚本。在应用任何其他自动脚本之前，应用于所有新查看器对象的可选脚本。指定用作基础自动脚本的脚本文件，以及其可执行程序将用于运行脚本的语言。

要应用自动脚本到输出项

- 在“命令标识符”网格中，选择一条命令以生成将对其应用自动脚本的输出项。

  在“对象和脚本”网格的对象列中，显示与选定命令关联的对象列表。脚本列显示选定命令的任何现有脚本。

- 为对象列中显示的任何项指定脚本。单击相应的脚本单元格。输入脚本路径，或单击省略号（...）按钮以浏览脚本。
- 指定其可执行程序将用于运行脚本的语言。注意：选定语言不受更改默认脚本语言的影响。
- 单击**应用**或**确定**。

删除自动脚本关联

- 在“对象和脚本”网格中，单击“脚本”列中与要取消关联的脚本对应的单元格。
- 删除脚本路径，然后单击“对象和脚本”网格中的任何其他单元格。
- 单击**应用**或**确定**。

语法编辑器选项

图片 17-11
“选项”对话框：“语法编辑器”选项卡



语法颜色编码

您可以关闭或打开命令、子命令、关键字、关键字值和注释的颜色编码，您可以指定每个的字体样式和颜色。

错误颜色编码

您可以关闭或打开某个语法错误的颜色编码，您可以指定使用的字体样式和颜色。包含错误的命令名称和文本（在命令内）都经过颜色编码，您可以选择每个的不同样式。

自动完成设置

您可以关闭或打开自动完成控制的自动显示。始终可以通过按 Ctrl+空格键显示自动完成控制。

缩进大小

指定缩进中的空格数量。该设置适用于缩进选定的语法行以及自动缩进。

装订线

这些选项指定在语法编辑器的装订线（为行号、书签、分界点和命令跨度保留的文本窗格左侧区域）内显示或隐藏行号和命令跨度的默认行为。命令跨度是提供命令开始和结束的可视指示符的图标。

窗格

显示导航窗格。 这指定显示或隐藏导航窗格的默认值。导航窗格在语法窗口中包含所有已识别命令的列表，以它们发生的顺序显示。单击导航窗格中的命令会将光标置于命令开始。

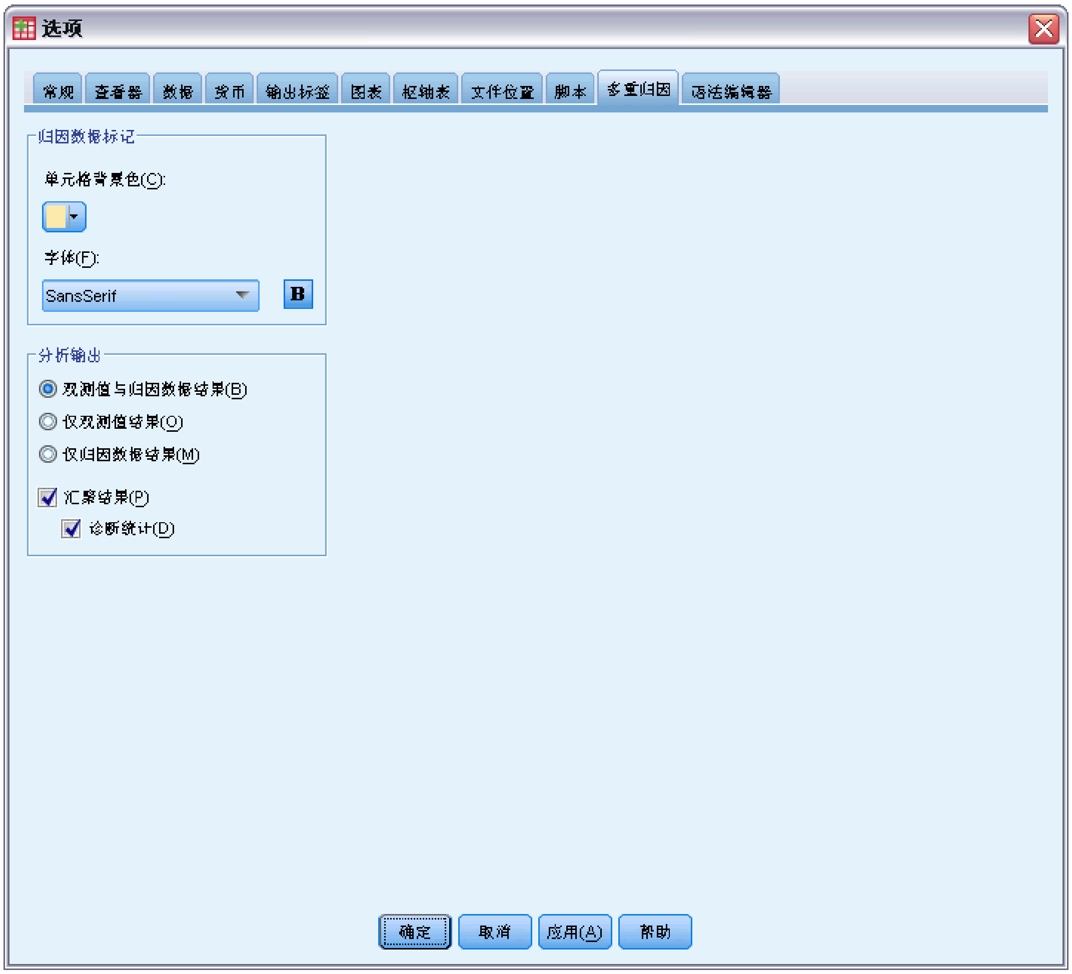
当发现错误时，自动打开“错误跟踪”窗格。 这指定当发现运行时间错误时显示或隐藏错误跟踪窗格的默认值。

针对从右至左语言优化。使用从右至左语言时，选中此复选框可获得最佳用户体验。

从对话框粘贴语法。 指定在从对话框粘贴语法时，语法在指定语法窗口中的插入位置。**在最后一个命令之后**会在最后一个命令之后插入粘贴的语法。**在光标位置或所选内容处**会在光标所在位置插入粘贴的语法；或者，如果已选择语法块，那么将用粘贴的语法替换所选语法块。

多重插补选项

图片 17-12
“选项”对话框：“多重插补”选项卡



“多重插补”选项卡控制与多重插补相关的两类首选项：

插补数据外观。缺省情况下，包含插补数据的单元格与包含非插补数据的单元格具有不同的背景颜色。插补数据的直观显示有助于您滚动数据集并找到这些单元格。还可以更改缺省单元格背景颜色、字体，以及使插补数据粗体显示等。

分析输出。这组首选项控制在分析多重插补数据集时产生的浏览器输出类型。缺省情况下，将为每个原始（插补前）数据集和插补数据集产生输出。此外，还为那些支持插补数据汇聚的过程生成最终的汇聚结果。当执行单变量汇聚时，还会显示汇聚诊断结果。不过，您可以隐藏那些不愿看到的输出。

设置多重插补选项

从菜单中选择：

编辑 > 选项

单击“多重插补”选项卡。

章

18

定制菜单和工具栏

菜单编辑器

可以使用菜单编辑器定制您的菜单。通过菜单编辑器，您可以：

- 添加运行定制脚本的菜单项。
- 添加运行命令语法文件的菜单项。
- 添加启动其他应用程序并自动将数据发送到其他应用程序的菜单项。

您能以以下格式将数据发送到其他应用程序：IBM® SPSS® Statistics、Excel、Lotus 1-2-3、制表符分隔格式以及 dBASE IV。

向菜单添加项

▶ 从菜单中选择：
视图 > **菜单编辑器**...

▶ 在菜单编辑器对话框中，双击想要对其添加新项的菜单（或单击加号图标）。

▶ 选择要在其上显示新项的菜单项。

▶ 单击**插入项**以插入新的菜单项。

▶ 为新项选择文件类型（脚本文件、命令语法文件或外部应用程序）。

▶ 单击**浏览**以选择要附加到菜单项的文件。

图片 18-1
菜单编辑器对话框



您还可以在菜单项之间添加全新的菜单和分隔符。另外，当您在菜单上选择另一个应用程序时，还可以自动将数据编辑器的内容发送到该应用程序。

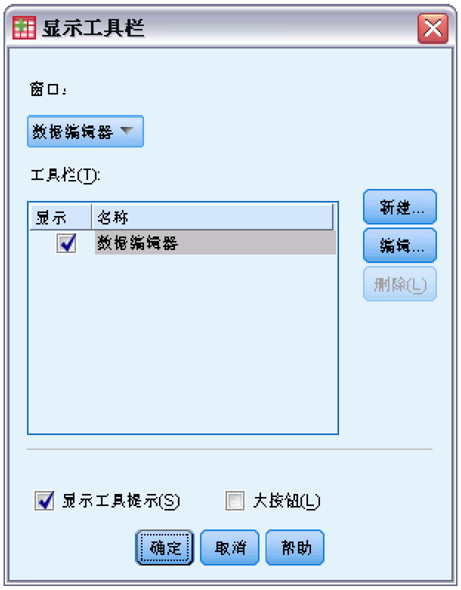
定制工具栏

您可以定制工具栏并创建新的工具栏。工具栏可以包含任何可用的工具，包括所有菜单操作的工具。它们还可以包含启动其他应用程序、运行命令语法文件或运行脚本文件的定制工具。

显示工具栏

使用“显示工具栏”可显示或隐藏工具栏、定制工具栏和创建新工具栏。工具栏可以包含任何可用的工具，包括所有菜单操作的工具。它们还可以包含启动其他应用程序、运行命令语法文件或运行脚本文件的定制工具。

图片 18-2
“显示工具栏”对话框



定制工具栏

- 从菜单中选择：
  视图 > 工具栏 > 自定义
- 选择想要定制的工具栏并单击**编辑**，或单击**新建**创建新的工具栏。
- 对于新的工具栏，请为工具栏输入名称，选择您想让工具栏出现的窗口，并单击**编辑**。
- 在“类别”列表中选择一项以显示该类别中的可用工具。
- 将需要的工具拖放到对话框中显示的工具栏上。
- 要从工具栏移去工具，则将其从对话框中显示的工具栏中拖走。

创建打开文件、运行命令语法文件或运行脚本的定制工具：

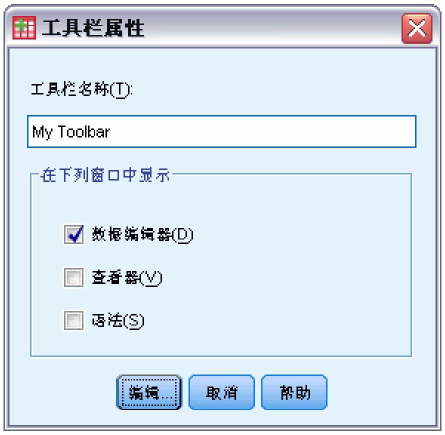
- 在“编辑工具栏”对话框中单击**新工具**。
- 为工具输入描述性标签。
- 选择需要该工具执行的操作（打开文件、运行命令语法文件或运行脚本）。
- 单击**浏览**选择要与此工具相关联的文件或应用程序。

新工具显示在用户定义的类别中，该类别还包含用户定义的菜单项。

工具栏属性

使用“工具栏属性”可选择想让选定的工具栏出现在其中的窗口类型。此对话框还用于为新工具栏创建名称。

图片 18-3
“工具栏属性”对话框



设置工具栏属性

▶ 从菜单中选择：
视图 > 工具栏 > 自定义

▶ 对于现有工具栏，单击**编辑**，然后在“编辑工具栏”对话框中单击**属性**。

▶ 对于新的工具栏，单击**新工具**。

▶ 选择想要让工具栏出现在其中的窗口类型。对于新的工具栏，还请输入工具栏名称。

编辑工具栏

使用“编辑工具栏”对话框可定制现有工具栏和创建新工具栏。工具栏可以包含任何可用的工具，包括所有菜单操作的工具。它们还可以包含启动其他应用程序、运行命令语法文件或运行脚本文件的定制工具。

图片 18-4
“定制工具栏”对话框



更改工具栏图像

▶ 选择想要更改其在工具栏上显示图像的工具。

▶ 单击**更改图像**。

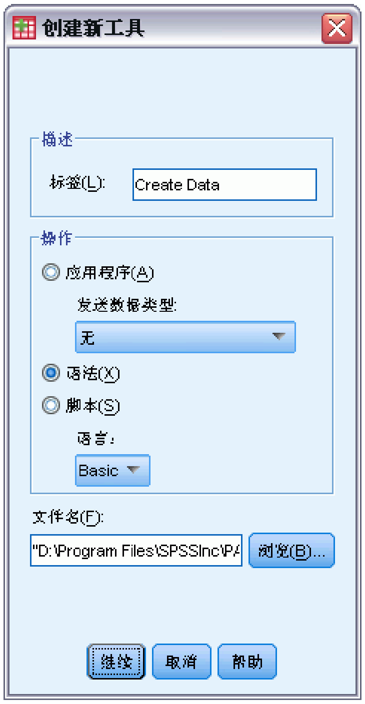
▶ 选择要用于工具的图像文件。支持下列图像格式：BMP、PNG、GIF 和 JPG。

- 图像必须为正方形。否则，图像会被切割为正方形形状。
- 图像会被自动缩放以适应要求。为获得最佳显示效果，请使用 16x16 像素图像作为小工具栏图像，或使用 32x32 像素图像作为大工具栏图像。

创建新工具

使用“创建新工具”对话框可创建启动其他应用程序、运行命令语法文件和运行脚本文件的定制工具。

图片 18-5
“创建新工具”对话框



章

19

创建和管理定制对话框

使用“定制对话框生成器”可以创建并管理定制对话框，以生成命令语法。使用“定制对话框生成器”，您可以：

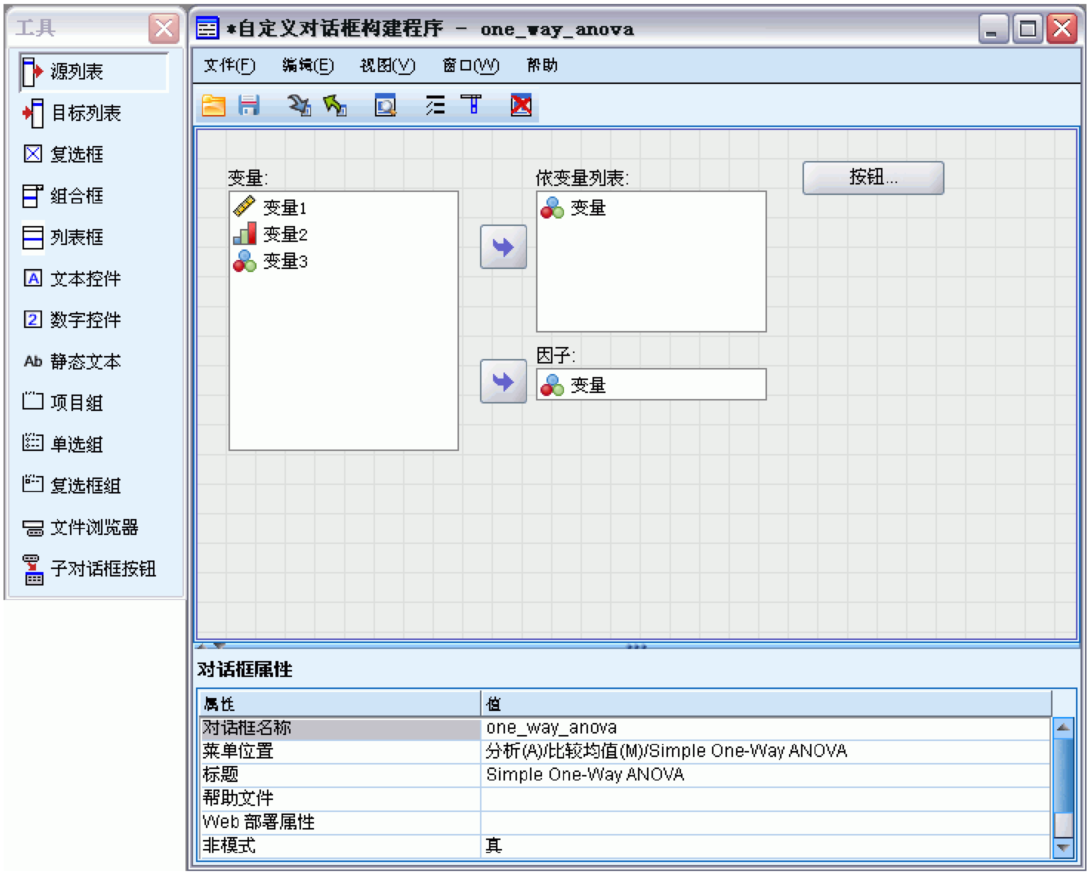
- 为内置 IBM® SPSS® Statistics 过程创建您个人版本的对话框。例如，可以为“频率”过程创建对话框，以允许用户选择变量集，然后通过预设的输出标准化选项产生命令语法。
- 创建用户界面以便为扩展命令生成命令语法。扩展命令为用户定义的 SPSS Statistics 命令，采用 Python 编程语言或 R 加以实现。
- 打开包含某个定制对话框（可能由其它用户创建）规范的文件，为您安装的 SPSS Statistics 添加对话框，还可进行修改。
- 保存定制对话框的规范，这样其他用户可以将其添加到他们安装的 SPSS Statistics 中。

如何启动定制对话框生成器

▶ 从菜单中选择：

实用程序 > 自定义对话框 > 定制对话框生成器...

定制对话框生成器布局



画布

画布是“定制对话框生成器”中设计对话框布局的区域。

属性窗格

属性窗格是“定制对话框生成器”中指定构成对话框的控件属性的区域，还可从中指定对话框本身的属性，例如菜单位置等。

工具调色板

工具调色板提供了一组可包含在定制对话框中的控件。您可以从“视图”菜单中选择“工具调色板”，以显示或隐藏工具调色板。

构建定制对话框

构建定制对话框的基本步骤包括：

- 指定对话框本身的属性，例如当打开对话框时显示的标题，以及对话框的新菜单项在 IBM® SPSS® Statistics 菜单中的位置。
- 指定构成对话框和子对话框的控件，例如源和目标变量列表。
- 创建语法模板以指定由对话框产生的命令语法。
- 为 SPSS Statistics 安装对话框和/或保存对话框规范至定制对话框包装文件 (.spd)。

在构建对话框时可以进行预览。

对话框属性

查看和设置“对话框属性”：

- 单击任何控件外部区域中的画布。如果画布上没有控件，“对话框属性”将始终可见。

对话框名称。“对话框名称”为必需属性，它指定了与对话框关联的唯一名称。在安装或卸载对话框时，使用该名称来标识对话框。为降低名称冲突可能性，您可能需要使用您组织的标识（例如 URL）来作为对话框名称的前缀。

菜单位置。单击省略号 (...) 按钮以打开菜单位置对话框，可以从中指定定制对话框的菜单项名称与位置。

标题。“标题”属性指定在对话框的标题栏中显示的文本。

帮助文件。“帮助文件”为可选属性，它指定了对话框帮助文件的路径。当用户单击对话框上的**帮助**按钮时，将打开此文件。帮助文件必须为 HTML 格式。当安装对话框或将其保存到定制对话框包装文件时，对话框规范连同指定帮助文件的副本一起包括在内。如果没有关联的帮助文件，运行时间对话框的帮助按钮将被隐藏。

一旦安装定制对话框之后，任何支持文件，例如图像文件与样式表，都必须与主帮助文件存储在一起。缺省情况下，对于 Windows 和 Linux，已安装定制对话框的规范存放在安装目录的 ext/lib/<对话框名称>文件夹中。对于 Mac，规范则存放在 /Library/Application Support/IBM/SPSS/Statistics/<version>/CustomDialogs/<对话框名称> 文件夹下，其中 <version> 为两位数的 IBM® SPSS® Statistics 版本，例如 19。支持文件应当位于文件夹的根目录而非子文件夹下。它们必须手动添加到您为对话框创建的定制对话框包装文件中。

如果为已安装对话框指定了备用位置（使用 SPSS_CDIALOGS_PATH 环境变量），则应在相应位置的<对话框名称>下存储所有支持文件。

注意：在使用从定制对话框包装文件 (.spd) 打开的对话框时，“帮助文件”属性指向与 .spd 文件关联的临时文件夹。任何对帮助文件的修改也必须应用到临时文件夹中的副本上。

Web 部署属性。允许您将属性文件关联到此对话框，以便在构建采用 Web 方式部署的瘦客户端应用程序时使用。

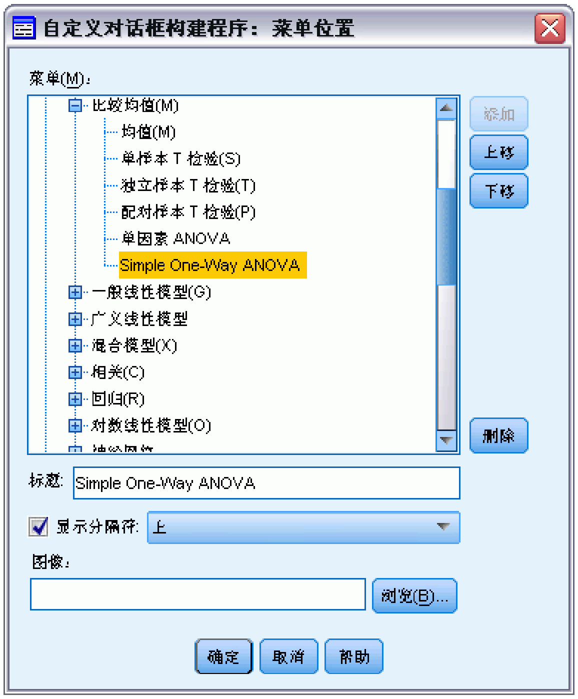
无模式。指定对话框为模式还是无模式对话框。如果为模式对话框，则它必须关闭后用户方可与主应用程序窗口（“数据”、“输出”和“语法”）或其它打开的对话框进行交互作用。无模式对话框没有此限制。缺省为无模式。

语法。“语法”属性指定语法模板，该模板用于创建对话框在运行时间生成的命令语法。单击省略号 (...) 按钮可打开“语法模板”。

所需插件。 指定一个或多个运行对话框生成的命令语法所需的插件，例如 Integration Plug-In for Python 或 Integration Plug-In for R。例如，如果对话框生成以 R 语言实现的扩展命令的命令语法，则选中 Integration Plug-In for R 对应的复选框。用户在安装或运行对话框时将得到缺少所需插件的警告消息。

为定制对话框指定菜单位置

图片 19-1
“定制对话框生成器菜单位置”对话框



使用“菜单位置”对话框可以为定制对话框指定菜单项的名称与位置。定制对话框的菜单项不会出现在 IBM® SPSS® Statistics 的“菜单编辑器”中。

▶ 双击想要对其添加新项的新对话框菜单（或单击加号图标）。还可以向标记为“定制”的顶级菜单（位于“图形”与“实用程序”项之间）添加菜单项，这仅对与定制对话框关联的菜单项可用。

如果要创建定制菜单或子菜单，请使用“菜单编辑器”。但请注意，该对话框的其他用户将必须从其“菜单编辑器”中手动创建相同的菜单或子菜单；否则，对话框将被添加到其“定制”菜单中。

注意：“菜单位置”对话框显示所有菜单，包括所有附加模块的菜单。确保将您定制对话框的菜单项添加到对您或对话框的最终用户可用的菜单。

▶ 选择要在其上显示新项的新对话框菜单项。在添加新项之后，可以使用**上移**和**下移**按钮调整其位置。

▶ 为菜单项输入标题。给定菜单或子菜单中的标题必须唯一。

▶ 单击**添加**。

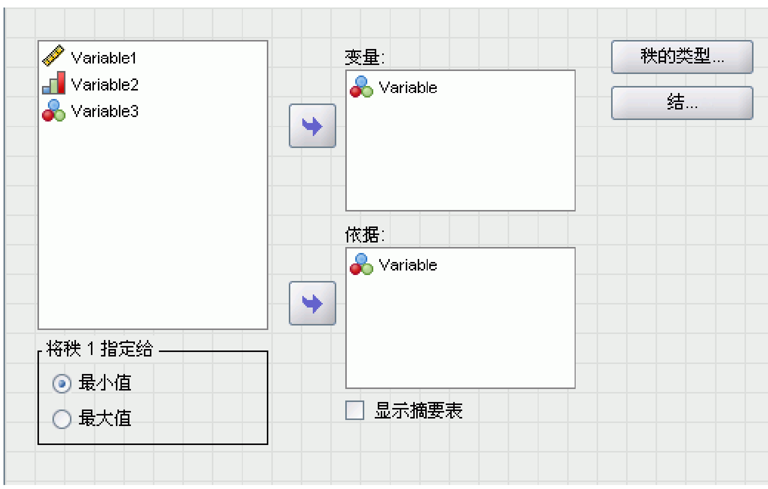
根据需要，您可以：

- 在新菜单项上面或下面添加分隔符。
- 为定制对话框指定其菜单项旁边所出现图像的路径。支持的图像格式为 gif 和 png。图像不得大于 16 x 16 像素。

在画布上布置控件

将控件从工具调色板拖到画布上，从而将其添加到定制对话框中。为确保与内置对话框的一致性，画布被划分为三个功能列，您可以在上面放入控件。

图片 19-2
画布结构



左列。左列主要供源列表控件使用。除目标列表与子对话框按钮以外的所有控件均可放入左列中。

中间列。中间列可以包含除源列表与子对话框按钮以外的所有控件。

右列。右列只可以包含子对话框按钮。

尽管未在画布上显示，每个定制对话框均包含有**确定**、**粘贴**、**取消**和**帮助**按钮，它们位于对话框的底部。这些按钮的存在及其位置均为自动确定。但如果没有帮助文件关联到对话框（由“对话框属性”中的“帮助文件”属性指定），则**帮助**按钮将被隐藏。

您可以通过将其向上或向下拖动更改一个列中的控件垂直顺序，但是控件的确切位置将为您自动确定。在运行期间，如果调整对话框本身的大小，控件将以相应的方式调整大小。诸如源和目标列表这类控件会自动扩展以填充其下方的可用空间。

构建语法模板

语法模板指定将由定制对话框产生的命令语法。单个定制对话框可为任意数量的内置 IBM® SPSS® Statistics 命令或扩展命令生成命令语法。

语法模板可以包含需要始终生成的静态文本，以及在运行时间被替换为相关定制对话框控件值的控件标识符。例如，不依赖于用户输入的命令名称和子命令指定可以作为静态文本，而在目标列表中指定的变量集则可使用目标列表控件的控件标识符来表示。

构建语法模板

▶ 从“自定义对话框生成器”的菜单中选择：
编辑 > 语法模板

（或在“对话框属性”的“语法”属性中单击省略号（...）按钮）

▶ 对于不依赖于用户指定值的静态命令语法，请按照?语法编辑器?中那样输入语法。“语法模板”对话框支持语法编辑器的自动完成和颜色编码功能。

▶ 在需要插入控件所生成命令语法的位置，以 %%Identifier%% 形式插入控件标识符，其中 Identifier 为控件的“标识符”属性的值。通过按 Ctrl+空格键，可以从可用控件标识符列表中进行选择。该列表包含的控件标识符后接可以使用语法自动完成功能的项。如果您手动输入标识符，请保留任何空格，因为标识符中的任何空格均有意义。

在运行期间，对于除复选框和复选框组以外的所有控件，每个标识符均被替换为相关控件中“语法”属性的当前值。对于复选框和复选框组，标识符将被替换为相关控件中“选中语法”或“未选中语法”属性的当前值，具体取决于控件的当前状态是选中还是未选中。

注意：如果没有终止符，在运行期间生成的语法将自动包括命令终止符（句点）作为最后一个字符。

示例:在语法模板中包含运行时间值

考虑“频率”对话框的简化版本，即只包含源列表控件与目标列表控件，并产生以下形式的命令语法：

```
FREQUENCIES VARIABLES=var1 var2...
/FORMAT = NOTABLE
 /BARCHART.
```

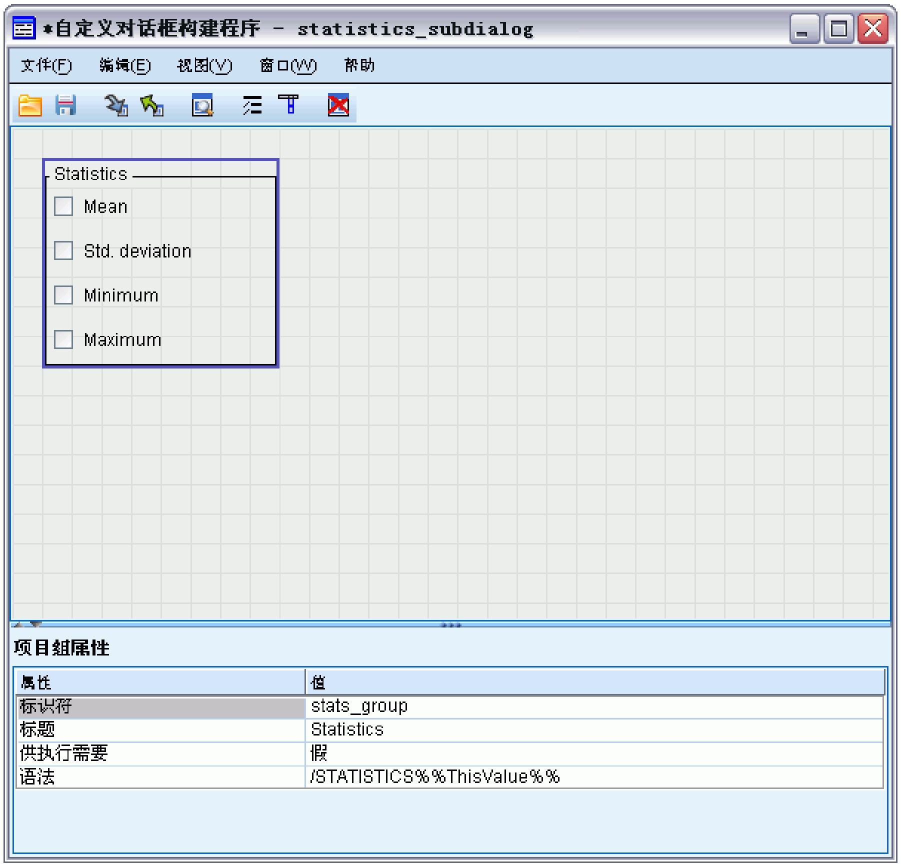
生成此命令语法的语法模板应形如：

```
FREQUENCIES VARIABLES=%%target_list%%
 /FORMAT = NOTABLE
 /BARCHART.
```

- %%target_list%% 为目标列表控件的“标识符”属性值。在运行期间，它将被替换为控件的“语法”属性的当前值。
- 将目标列表控件的“语法”属性定义为 %%ThisValue%%，如此一来在运行期间，该属性的当前值将指定为控件的值，即目标列表中的变量集。

示例:包含来自容器控件的命令语法

在前一个示例基础上，考虑添加一个包含单组复选框的“统计量”子对话框，以允许用户指定均值、标准差、最小值和最大值。假定复选框包含在项组控件中，如下图所示。



生成的命令语法示例为：

```
FREQUENCIES VARIABLES=var1 var2...
/FORMAT = NOTABLE
  /STATISTICS MEAN STDDEV
  /BARCHART.
```

生成此命令语法的语法模板应形如：

```
FREQUENCIES VARIABLES=%%target_list%%
  /FORMAT = NOTABLE
  %%stats_group%%
  /BARCHART.
```

- %%target_list%% 为目标列表控件的“标识符”属性值，而 %%stats_group%% 则为项组控件的“标识符”属性值。

下表所示为一种方式，能通过指定项组及其中所包含复选框的“语法”属性以生成所需结果。如前一个示例中所述，目标列表的“语法”属性将被设为 %%ThisValue%%。

项组的“语法”属性	/STATISTICS %%ThisValue%%
均值复选框的“选中语法”属性	MEAN
标准差复选框的“选中语法”属性	STDDEV

最小值复选框的“选中语法”属性	MINIMUM
最大值复选框的“选中语法”属性	MAXIMUM

在运行期间，%%stats_group%% 将被替换为项组控件的“语法”属性的当前值。具体地，%%ThisValue%% 将被替换为每个复选框的“选中语法”或“未选中语法”属性（取决于其状态为选中还是未选中）的值所组成的空格分隔列表。由于只为?选中语法?属性指定了值，因此只有选中的复选框才在 %%ThisValue%% 中出现。例如，如果用户选中了均值和标准差复选框，则项组控件的?语法?属性的运行时值将为 /STATISTICS MEAN STDDEV。

如果未选中任何复选框，则项组控件的?语法?属性将为空，生成的命令语法也不会包含任何对 %%stats_group%% 的引用。这可能符合也可能不符合需要。例如，即使未选中任何复选框，您可能仍然希望生成 STATISTICS 子命令。这可以通过在语法模板中直接引用复选框的标识符来实现，即：

```
FREQUENCIES VARIABLES=%%target_list%%
 /FORMAT = NOTABLE
 /STATISTICS %%stats_mean%% %%stats_stddev%% %%stats_min%% %%stats_max%%
 /BARCHART.
```

预览定制对话框

可以预览当前在“定制对话框生成器”中打开的对话框。这将显示对话框，其功能与从 IBM® SPSS® Statistics 菜单中运行它时的功能相同。

- 源变量列表中被填入虚变量，并可将其移动到目标列表。
- **粘贴**按钮将命令语法粘贴到指定的语法窗口中。
- 单击**确定**可关闭预览。
- 如果指定了帮助文件，则会启用**帮助**按钮，并可用它来打开指定文件。如果未指定帮助文件，则预览时会禁用帮助按钮；当实际对话框运行时，此按钮也将被隐藏。

预览定制对话框：

▶ 从“自定义对话框生成器”的菜单中选择：
文件 > 预览对话框

管理定制对话框

使用“定制对话框生成器”可以管理您或其他用户创建的定制对话框。可以安装、卸载或修改已安装的对话框；还可以将定制对话框的规范保存到外部文件，或打开包含定制对话框规范的文件并进行修改。在使用定制对话框之前必须先进行安装。

安装定制对话框

可以安装当前在“定制对话框生成器”中打开的对话框，或从定制对话框包装文件 (.spd) 安装对话框。重新安装存在的对话框将替换已有版本。

安装当前打开的对话框：

- 从“自定义对话框生成器”的菜单中选择：
  文件 > 安装

从定制对话框包装文件安装：

- 从菜单中选择：
  实用程序 > 自定义对话框 > 安装定制对话框...

对于 Windows 和 Linux，默认情况下，安装对话框需要 IBM® SPSS® Statistics 安装目录的写入权限（对于 Mac，对话框安装到一般用户可写入的位置）。如果对所需位置没有写入权限，或打算在其它位置保存安装的对话框，则可通过定义 SPSS_CDIALOGS_PATH 环境变量以指定一个或多个备用位置。如果在 SPSS_CDIALOGS_PATH 中指定了路径，则它优先于缺省位置。定制对话框将被安装到第一个可写入位置。注意， Mac 用户还可以使用 SPSS_CDIALOGS_PATH 环境变量。如果要指定多个备用位置，对于 Windows 请使用分号分隔，对于 Linux 和 Mac，则使用冒号分隔。指定位置必须在目标计算机上存在。设置 SPSS_CDIALOGS_PATH 后，您需要重新启动 SPSS Statistics 以使更改生效。

要在 Windows 上创建 SPSS_CDIALOGS_PATH 环境变量，从控制面板中：

Windows XP

- 选择“系统”。
- 选择“高级”选项卡并单击**环境变量**。
- 在“用户变量”部分，单击**新建**，在变量名字段中输入 SPSS_CDIALOGS_PATH，并在变量值字段中输入所需路径。

Windows Vista 或 Windows 7

- 选择“用户帐户”。
- 选择“更改我的环境变量”。
- 单击**新建**，在“变量名”字段中输入 SPSS_CDIALOGS_PATH ，并在“变量值”字段中输入所需路径。

要查看定制对话框的当前位置，请运行下列命令语法：SHOW EXTPATHS。

打开已安装的定制对话框

可以打开已安装的定制对话框，并进行修改和/或保存到外部文件，以便分发给其他用户。

- 从“自定义对话框生成器”的菜单中选择：
  文件 > 打开已安装的

注意：如果打开了已安装的对话框并进行修改，选择“文件 > 安装”将重新安装对话框，并替换现有版本。

卸载定制对话框

▶ 从“自定义对话框生成器”的菜单中选择：
文件 > 卸载

下次启动 SPSS Statistics 时，定制对话框的菜单项将被删除。

保存到定制对话框包装文件

可以将定制对话框的规范保存到外部文件，这样就可以向其他用户分发对话框，或保存那些尚未安装的对话框的规范。规范被保存到定制对话框包装文件（.spd）。

▶ 从“自定义对话框生成器”的菜单中选择：
文件 > 保存

打开定制对话框包装文件

可以打开包含定制对话框规范的定制对话框包装文件，以便进行修改并重新保存或安装对话框。

▶ 从“自定义对话框生成器”的菜单中选择：
文件 > 打开

手动复制已安装的定制对话框

缺省情况下，对于 Windows 和 Linux，已安装定制对话框的规范存放在安装目录的 ext/lib/<对话框名称>文件夹中。对于 Mac，规范则存放在 /Library/Application Support/IBM/SPSS/Statistics/<version>/CustomDialogs/<对话框名称> 文件夹下，其中 <version> 为两位数的 SPSS Statistics 版本，例如 19。您可以将此文件夹复制到另一个 SPSS Statistics 实例中的相同位置，在下次启动该实例时，会将其识别为已安装的对话框。

- 如果为已安装对话框指定了备用位置（使用 SPSS_CDIALOGS_PATH 环境变量），则应复制到相应位置的<对话框名称>文件夹下。
- 如果为 SPSS Statistics 实例定义了已安装对话框的备用位置，则可以复制到指定的位置之一，下次启动该实例时，会将其识别为已安装的对话框。

控件类型

工具调色板提供了可添加到定制对话框中的控件。

- **源列表**：活动数据集中源变量的列表。
- **目标列表**:从源列表中转移变量的目标。
- **复选框**：单个复选框。
- **组合框**:用于创建下拉列表的组合框。
- **列表框**: 用于创建单选或多选列表的列表框。
- **文本控件**：接受任意文本作为输入的文本框。
- **数字控件**：仅接受数值作为输入的文本框。

- **静态文本控件**：用于显示静态文本的控件。
- **项组**：包含一系列控件（如一组复选框）的容器。
- **单选按钮组**：一组单选按钮。
- **复选框组**：包含一组通过单个复选框进行启用或禁用的控件的容器。
- **文件浏览器**：用于浏览文件以打开或保存文件的控件。
- **子对话框按钮**：用于打开子对话框的按钮。

源列表

“源变量列表”控件显示来自活动数据集，且对对话框的最终用户可用的变量列表。可以显示所有来自活动数据集的变量（缺省情况下），或根据类型和测量级别来过滤列表，例如，具有刻度测量级别的数值变量。使用源列表控件，表明将使用一个或多个目标列表控件。源列表控件具有下列属性：

标识符。 控件的唯一标识。

标题。 出现在控件上方的可选标题。对于多行标题，可使用 \n 指定换行。

工具提示。当用户悬停在控件上时，显示的可选工具提示文本。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

变量转移。指定从源列表转移到目标列表的变量是保留在源列表中（复制变量），还是从源列表中删除（移动变量）。

变量过滤。可以过滤在控件中显示的变量集。可以按类型与测量级别过滤，并可指定在变量列表中包括多重响应集。单击省略号 (...) 按钮可打开“过滤”对话框。还可通过双击画布上的“源列表”控件以打开“过滤”对话框。

注意：源列表控件不能添加到子对话框中。

目标列表

源列表控件提供从源列表中转移变量的目标。使用目标列表控件，假设存在相应的源列表控件。可以指定仅单个变量能够被转移到控件，或能够转移多个变量；还可以限制转移到控件的变量类型，例如，只转移具有名义或有序测量级别的数值变量。目标列表控件具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。 出现在控件上方的可选标题。对于多行标题，可使用 \n 指定换行。

工具提示。当用户悬停在控件上时，显示的可选工具提示文本。

目标列表类型。指定可以转移多个变量还是只可转移单个变量到控件。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

执行需要。指定该控件中是否需要值，以便继续执行操作。如果指定 True，则**确定**和**粘贴**按钮将被禁用，直到为该控制指定值为止。如果指定 False，该控件中缺少值对**确定**和**粘贴**按钮的状态没有效应。True。

变量过滤。可以限制能够转移到控件的变量类型。可以按类型与测量级别进行限制，并可指定是否能够转移多重响应集到控件。单击省略号 (...) 按钮可打开"过滤"对话框。还可通过双击画布上的"目标列表"控件以打开"过滤"对话框。

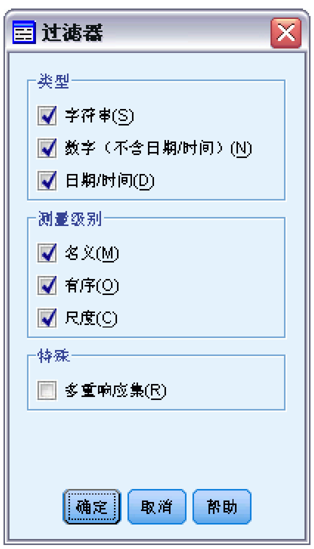
语法。 指定该控件在运行时间生成的命令语法，并可插入到语法模板中。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。
- 值 %%ThisValue%% 指定控件的运行时值，它是转移到控件的变量列表。这是默认值。

注意：目标列表控件不能添加到子对话框中。

过滤变量列表

图片 19-3
"过滤"对话框



"过滤"对话框与源和目标变量列表相关联，使用它可以过滤来自活动数据集并显示在列表中的变量类型。还可以指定是否包含与活动数据集关联的多重响应集。数值变量包括除日期与时间格式外的所有数值格式。

复选框

"复选框"控件为简单复选框，可以针对选中和未选中状态生成不同的命令语法。复选框控件具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。与复选框一起显示的标签。对于多行标题，可使用 \n 指定换行。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

缺省值。复选框的缺省状态为选中或未选中。

选中/未选中语法。指定在选中和未选中控件时生成的命令语法。要在语法模板中包含命令语法，请使用“标识符”属性的值。生成的语法，不论来自“选中语法”还是“未选中”语法，都会被插入标识符的指定位置。例如，如果标识符为 checkbox1，则在运行期间，语法模板中 %%checkbox1%% 的实例将在选中复选框时被替换为“选中语法”属性的值，在未选中复选框时被替换为“未选中语法”属性的值。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。

组合框和列表框控件

使用“组合框”控件可以创建下拉列表，以生成特定于所选列表项的命令语法。它限制为单选。使用“列表框”控件可以显示支持单选或多选的列表项，并可生成特定于所选项的命令语法。组合框和列表框控件具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。 出现在控件上方的可选标题。对于多行标题，可使用 \n 指定换行。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

列表项。单击省略号 (...) 按钮以打开列表项属性对话框，可以从中指定控件的列表项。还可通过双击画布上的“组合框”或“列表框”控件以打开“列表项属性”对话框。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

列表框类型（仅列表框）。指定列表框是仅支持单选，还是支持多选。还可指定列表项显示为复选框列表。

语法。 指定该控件在运行时间生成的命令语法，并可插入到语法模板中。

- 值 %%ThisValue%% 指定控件的运行时间值，它是缺省值。如果手动定义列表项，则运行时间值为选定列表项的“语法”属性值。如果列表项是基于目标列表控件，则运行时间值为选定列表项的值。对于多选列表框控件，运行时间值为以空格分隔的选定项列表。
- 您可指定任何有效的命令语法，并使用 \n 作为换行符。

为组合框和列表框指定列表项

使用“列表项属性”对话框可以指定组合框或列表框控件的列表项。

手动定义的值。 允许您明确指定每个列表项。

- **标识符。**列表项的唯一标识。
- **名称。**在列表中出现的该项的名称。名称为必需的字段。
- **缺省值。**对于组合框，指定列表项是否为组合框中显示的缺省项。对于列表框，指定列表项是否缺省选中。

- **语法。**指定在选中列表项时生成的命令语法。
- 您可指定任何有效的命令语法，并使用 \n 作为换行符。

注意：可以在现有列表底部的空白行中添加新的列表项。输入标识符以外的任何属性会产生唯一的标识符，您可以保留或更改此标识符。要删除列表项，请单击其标识符单元格，并按“删除”。

基于目标列表控件内容的值。指定使用选定目标列表控件中变量的关联值动态填充列表项。选择现有的目标列表控件作为列表项的源，或者在目标列表组合框的文本区域中输入目标列表控件的“标识符”属性值。使用后一种方法可以输入您打算稍后添加的目标列表控件的标识符。

- **变量名。**使用指定目标列表控件中的变量名填充列表项。
- **值标签。**使用指定目标列表控件中变量的关联值标签集合填充列表项。您可以选择由关联组合框或列表框控件生成的命令语法是否包含选定值标签或其相关值。
- **定制属性。**使用包含指定定制属性的目标列表控件中变量的关联属性值集合填充列表项。
- **语法。**显示关联组合框或列表框控件的“语法”属性，并允许您更改属性。

文本控件

“文本”控件为简单文本框，它可接受任意文本输入，并具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。 出现在控件上方的可选标题。对于多行标题，可使用 \n 指定换行。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

文本内容。指定内容为任意文本，或者必需包含符合 IBM® SPSS® Statistics 变量名规则的字符串。

缺省值。文本框的缺省内容。

执行需要。指定该控件中是否需要值，以便继续执行操作。如果指定 True，则**确定**和**粘贴**按钮将被禁用，直到为该控制指定值为止。如果指定 False，该控件中缺少值对**确定**和**粘贴**按钮的状态没有效应。False。

语法。 指定该控件在运行时间生成的命令语法，并可插入到语法模板中。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。
- 值 %%ThisValue%% 指定控件的运行时间值，它是文本框的内容。这是默认值。
- 如果语法属性包括 %%ThisValue%%，且文本框的运行时间值为空，则文本框控件不会生成任何命令语法。

数字控件

“数字”控件为文本框，它用于输入数值，并具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。 出现在控件上方的可选标题。对于多行标题，可使用 \n 指定换行。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

数值类型。指定对输入内容的任何限制。如果值为“Real”，则表示对输入值除了必须为数值外，没有其它限制。如果值为“Integer”，表示输入值必须为整数。

缺省值。缺省值，如果有的话。

最小值。最小允许值，如果有的话。

最大值。最大允许值，如果有的话。

执行需要。指定该控件中是否需要值，以便继续执行操作。如果指定 True，则**确定**和**粘贴**按钮将被禁用，直到为该控制指定值为止。如果指定 False，该控件中缺少值对**确定**和**粘贴**按钮的状态没有效应。False。

语法。 指定该控件在运行时间生成的命令语法，并可插入到语法模板中。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。
- 值 %%ThisValue%% 指定控件的运行时间值，它是数值。这是默认值。
- 如果语法属性包括 %%ThisValue%%，且数字控件的运行时间值为空，则数字控件不会生成任何命令语法。

静态文本控件

使用“静态文本”控件可为对话框添加一个文本块，它具有下列属性：

标识符。 控件的唯一标识。

标题。文本块的内容。对于多行内容，可使用 \n 指定换行。

项目组

“项组”控件是其它控件的容器，通过它可以对多个控件生成的语法进行分组和控制。例如，如果使用一组复选框指定某个子命令的可选设置，但只希望在选中至少一个复选框时生成子命令语法。这可通过使用“项组”控件作为复选框控件的容器来实现。在“项组”中可以包含下列控件类型：复选框、组合框、文本控件、数字控件、静态文本、单选按钮组和文件浏览器。项组控件具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。组的可选标题。对于多行标题，可使用 \n 指定换行。

执行需要。如果指定 True，则**确定**和**粘贴**按钮将被禁用，直到该组中至少一个控件具有值为止。缺省为 False。

例如，组包含一系列复选框。如果“执行需要”被设为 True 且所有复选框均未选中，则**确定**和**粘贴**将被禁用。

语法。 指定该控件在运行时间生成的命令语法，并可插入到语法模板中。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。
- 可以将标识符包含在“项组”中的任何控件中。在运行期间，标识符将被替换为控件生成的语法。
- 值 %%ThisValue%% 生成项组中每个控件所产生语法的空格分隔列表，并以它们在组中的显示顺序排列（从上到下）。这是默认值。如果语法属性包括 %%ThisValue%%，且项组中所有控件均未生成语法，则项组整体也不会生成任何命令语法。

单选组

“单选按钮组”控件是一组单选按钮的容器，每个单选按钮又可包含一组嵌套控件。单选按钮组控件具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。组的可选标题。如果略过，则不会显示组边框。对于多行标题，可使用 \n 指定换行。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

单选按钮。单击省略号（...）按钮以打开单选按钮组属性对话框，可以从中单选按钮的属性，并可为组添加或删除按钮。在给定单选按钮下嵌套控件的功能是单选按钮的属性，在“单选按钮组属性”对话框中进行设置。注意：还可通过双击画布上的“单选按钮组”控件以打开“单选按钮组属性”对话框。

语法。 指定该控件在运行时间生成的命令语法，并可插入到语法模板中。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。
- 值 %%ThisValue%% 指定单选按钮组的运行时间值，它是所选单选按钮的“语法”属性值。这是默认值。如果语法属性包括 %%ThisValue%%，且选定单选按钮未生成语法，则单选按钮组也不会生成任何命令语法。

定义单选按钮

使用“单选按钮组属性”对话框中可以指定一组单选按钮。

标识符。单选按钮的唯一标识。

名称。单选按钮旁边显示的名称。名称为必需的字段。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

记忆键。在名称中用作记忆键的可选字符。指定字符必须存在于名称中。

嵌套组。指定是否可以在此单选按钮下嵌套其他控件。缺省为 False。如果嵌套组属性设置为 True，则会在相关单选按钮下显示矩形放置区，并且可以嵌套和缩进。下列控件可以嵌套在单选按钮下：复选框、文本、静态文本、数字、组合框、列表框和文件浏览器。

缺省值。指定单选按钮是否缺省选中。

语法。指定在选中单选按钮时生成的命令语法。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。
- 对于包含嵌套控件的单选按钮，值 %%ThisValue%% 生成每个嵌套控件所产生语法的空格分隔列表，并以它们在单选按钮下的显示顺序排列（从上到下）。

可以在现有列表底部的空白行中添加新的单选按钮。输入标识符以外的任何属性会产生唯一的标识符，您可以保留或更改此标识符。要删除单选按钮，请单击其标识符单元格，并按“删除”。

复选框组

“复选框组”控件是包含一组通过单个复选框进行启用或禁用的控件的容器。在“复选框组”中可以包含下列控件类型：复选框、组合框、文本控件、数字控件、静态文本、单选按钮组和文件浏览器。复选框组控件具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。组的可选标题。如果略过，则不会显示组边框。对于多行标题，可使用 \n 指定换行。

复选框标题。与控制复选框一起显示的可选标签。对于多行标题，可使用 \n 指定换行。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

缺省值。控制复选框的缺省状态为选中或未选中。

选中/未选中语法。指定在选中和未选中控件时生成的命令语法。要在语法模板中包含命令语法，请使用“标识符”属性的值。生成的语法，不论来自“选中语法”还是“未选中”语法，都会被插入标识符的指定位置。例如，如果标识符为 checkboxgroup1，则在运行期间，语法模板中 %%checkboxgroup1%% 的实例将在选中复选框时被替换为“选中语法”属性的值，在未选中复选框时被替换为“未选中语法”属性的值。

- 您可指定任何有效的命令语法，并使用 \n 作为换行符。
- 可以将标识符包含在“复选框组”中的任何控件中。在运行期间，标识符将被替换为控件生成的语法。
- 值 %%ThisValue%% 可以用在“选中语法”或“未选中语法”属性中。它可生成复选框组中每个控件所产生语法的空格分隔列表，并以它们在组中的显示顺序排列（从上到下）。
- 缺省情况下，“选中语法”属性的值为 %%ThisValue%%，“未选中语法”属性为空值。

文件浏览器

“文件浏览器”控件包含一个文件路径文本框和一个浏览器按钮，它可打开一个标准 IBM® SPSS® Statistics 对话框以打开或保存文件。文件浏览器控件具有下列属性：

标识符。 控件的唯一标识。这是在语法模板中引用控件时使用的标识符。

标题。 出现在控件上方的可选标题。对于多行标题，可使用 \n 指定换行。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

文件系统操作。指定由浏览器按钮打开的对话框是用来打开文件还是保存文件。如果值为“Open”，则表示浏览器对话框将验证指定文件是否存在。如果值为“Save”，表示浏览器对话框不会验证指定文件是否存在。

浏览器类型。指定浏览器对话框是用来选择文件（定位文件）还是选择文件夹（定位文件夹）。

文件过滤。单击省略号（...）按钮以打开文件过滤对话框，可以从中指定打开或保存对话框的可用文件类型。缺省情况下允许所有文件类型。注意：还可通过双击画布上的“文件浏览器”控件以打开“文件过滤”对话框。

文件系统类型。在分布式分析模式下，这可以指定打开或保存对话框是浏览 SPSS Statistics 服务器运行所在的文件系统，还是本地计算机的文件系统。选择**服务器**浏览服务器的文件系统，或选择**客户端**浏览本地计算机的文件系统。该属性对本地分析模式没有影响。

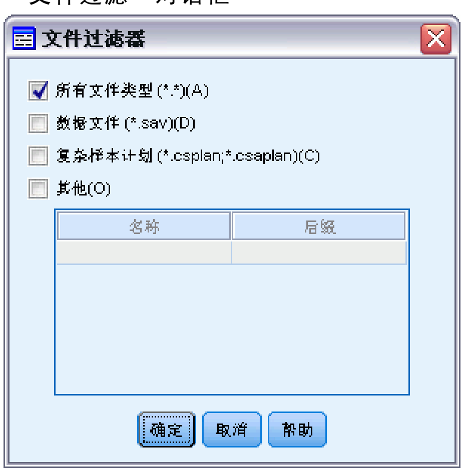
执行需要。指定该控件中是否需要值，以便继续执行操作。如果指定 `True`，则**确定**和**粘贴**按钮将被禁用，直到为该控制指定值为止。如果指定 `False`，该控件中缺少值对**确定**和**粘贴**按钮的状态没有效应。`False`。

语法。 指定该控件在运行时间生成的命令语法，并可插入到语法模板中。

- 您可指定任何有效的命令语法，并使用 `\n` 作为换行符。
- 值 `%%ThisValue%%` 指定文本框的运行时间值，它是手动或由对话框输入的文件路径。这是默认值。
- 如果语法属性包括 `%%ThisValue%%`，且文本框的运行时间值为空，则文件浏览器控件不会生成任何命令语法。

文件类型过滤

图片 19-4
“文件过滤”对话框



使用“文件过滤”对话框可以为从“文件系统浏览器”控件中访问的打开和保存对话框指定其“文件类型”和“另存类型”下拉列表中显示的文件类型。缺省情况下允许所有文件类型。

指定在对话框中不显式列出的文件类型：

- ▶ 选择“其它”。
- ▶ 输入文件类型名称。
- ▶ 以 *.**后缀名**形式输入文件类型，例如 *.xls。可以指定多个文件类型，并以分号隔开。

子对话框按钮

“子对话框按钮”指定用于打开子对话框的按钮，以便在对话框生成器中访问子对话框。子对话框按钮控件具有下列属性：

标识符。 控件的唯一标识。

标题。在按钮中显示的文本。

工具提示。 当用户悬停在控件上时，显示的可选工具提示文本。

子对话框。单击省略号（...）按钮可打开子对话框的“定制对话框生成器”。还可双击子对话框按钮，以打开生成器。

记忆键。 在标题中用作控件的键盘快捷键的可选字符。这些字符加下划线出现在标题中。按 Alt+[记忆键]可激活快捷方式。在 Mac 上不支持“记忆键”属性。

注意：“子对话框按钮”控件不能添加到子对话框。

子对话框的对话框属性

查看和设置子对话框属性：

- ▶ 在主对话框中双击子对话框按钮以打开子对话框，或单击子对话框按钮并单击省略号（...）按钮以显示对话框属性。
- ▶ 在子对话框中，单击任何控件外部区域中的画布。如果画布上没有控件，子对话框属性将始终可见。

子对话框名称。子对话框的唯一标识。“子对话框名称”为必需字段。

注意：如果在语法模板中指定“子对话框”作为标识符，即 %%My Sub-dialog Name%%，它将在运行时被替换为子对话框中每个控件所生成语法的空格分隔列表，并以其显示顺序排列（从上到下和从左到右）。

标题。指定在子对话框的标题栏中显示的文本。“标题”属性为可选，但建议填入值。

帮助文件。为子对话框指定可选帮助文件的路径。当用户单击子对话框上的**帮助**按钮时，将打开此文件，它可以是为主对话框指定的帮助文件。帮助文件必须为 HTML 格式。有关更新信息，请参见对话框属性的“帮助文件”属性说明。

语法。单击省略号（...）按钮可打开“语法模板”。

扩展命令的自定义对话框

扩展命令为用户定义的 IBM® SPSS® Statistics 命令，采用 Python 编程语言或 R 加以实现。一旦部署到 SPSS Statistics 实例之后，扩展命令的运行方式与内部 SPSS Statistics 命令相同。

可以使用“定制对话框生成器”来为扩展命令创建对话框，还可以安装其他用户创建的扩展命令定制对话框。

为扩展命令创建定制对话框

不论扩展命令是由您自己还是其他用户编写，都可以为其创建定制对话框。对话框的语法模板将为扩展命令生成命令语法。如果定制对话框仅供您使用，请安装此对话框。假设扩展命令已部署到您的系统上，这样您就可以从安装的对话框中运行该命令。

如果您为扩展命令创建定制对话框，并打算共享给其他用户，则需要首先将对话框规范保存到定制对话框包装文件 (.spd)。然后，您将需要创建一个扩展束，其中包括定制对话框包装文件，规定扩展命令语法的 XML 文件，以及以 Python 或 R 编写的实现代码文件。您随后可以将扩展束共享给其他用户。

安装带有相关定制对话框的扩展命令

带有相关定制对话框的扩展命令包含三个部分：一个规定命令语法的 XML 文件；实现命令的一个或多个代码文件（Python 或 R 语言）；一个包含定制对话框规范的定制对话框包装文件。如果扩展命令及其关联定制对话框以扩展束 (.spe) 文件形式分发，则您只需通过“实用程序>扩展束>安装扩展束”即可安装该束。否则，您需要以下列方式单独安装定制对话框和扩展命令文件：

定制对话框包装文件。 通过“实用程序>定制对话框>安装定制对话框”安装定制对话框包装文件。

XML 语法规范文件和实现代码。 对于 Windows 和 Linux，规定扩展命令语法的 XML 文件和实现代码（Python 或 R 语言）必须位于 SPSS Statistics 安装目录的 extensions 目录下。对于 Mac，XML 和代码文件必须位于 /Library/Application Support/IBM/SPSS/Statistics/19/extensions 目录。

- 对于 Windows 和 Linux，如果您对 SPSS Statistics 安装目录没有写入权限，或打算在其它位置保存 XML 文件和实现代码，则可通过定义 SPSS_EXTENSIONS_PATH 环境变量指定一个或多个备用位置。如果在 SPSS_EXTENSIONS_PATH 中指定了路径，则它优先于 SPSS Statistics 安装目录。注意，Mac 用户还可以使用 SPSS_EXTENSIONS_PATH 环境变量。如果要指定多个备用位置，对于 Windows 请使用分号分隔，对于 Linux 和 Mac，则使用冒号分隔。

  要在 Windows 上创建 SPSS_EXTENSIONS_PATH 环境变量，请遵循与创建 SPSS_CDIALOGS_PATH 变量相同的一般步骤。请参见管理定制对话框第 325 页码中的“安装定制对话框”部分。

  要查看定制对话框的当前位置，请运行下列命令语法：SHOW EXTPATHS。
- 对于以 Python 实现的扩展命令，始终需要将相关 Python 模块存储到被列入 Python 搜索路径的位置中，例如 Python site-packages 目录。

注意：要使用新的扩展命令，重新启动 SPSS Statistics。

创建定制对话框的本地化版本

可以针对 IBM® SPSS® Statistics 支持的任何语言创建定制对话框的本地化版本。定制对话框中显示的所有字符串，以及可选的帮助文件，均可进行本地化。

本地化对话框字符串

- ▶ 复制对话框相关的属性文件。属性文件包含了对话框中所有可本地化的字符串。默认情况下，对于 Windows 和 Linux，它位于 SPSS Statistics 安装目录的 ext/lib/<对话框名称> 文件夹中；对于 Mac，则位于 /Library/Application Support/IBM/SPSS/Statistics/19/CustomDialogs/<对话框名称> 文件夹中。该副本应位于同一文件夹而非子文件夹下。

  如果为已安装对话框指定了备用位置（使用 SPSS_CDIALOGS_PATH 环境变量），则副本应位于相应位置的<对话框名称>文件夹下。

- ▶ 将副本重命名为<对话框名称>_<语言标识符>.properties，请使用下表列出的语言标识。例如，如果对话框名称为 mydialog，且您打算创建日语版本的对话框，则本地化后的属性文件应命名为 mydialog_ja.properties。本地化属性文件必须手动添加到您为对话框创建的定制对话框包装文件中。定制对话框包装文件为简单的 ZIP 文件，可以通过诸如 Windows 上的 WinZip 等应用程序打开并修改。

- ▶ 使用支持 UTF-8 的文本编辑器打开新的属性文件，例如 Windows 上的 Notepad，或 Mac 上的 TextEdit 应用程序。修改需要本地化的任何属性的相关值，但不要更改属性名称。与特定控件关联的属性是以控件的标识符为前缀。例如，某个标识符为 options_button 的控件的“工具提示”属性为 options_button_tooltip_LABEL。“标题”属性则简单地命名为 <标识符>_LABEL，如 options_button_LABEL。

  当打开对话框时，SPSS Statistics 将搜索其语言标识符与当前语言匹配的属性文件，当前语言在“选项”对话框的“常规”选项卡上的”语言”下拉列表中指定。如果找不到该属性文件，则使用缺省文件<对话框名称>.properties。

本地化帮助文件

- ▶ 复制对话框相关的帮助文件，并将其文字本地化成所需语言。该副本应与帮助文件位于同一文件夹而非子文件夹下。对于 Windows 和 Linux，帮助文件位于 SPSS Statistics 安装目录的 ext/lib/<对话框名称> 文件夹中；对于 Mac，则位于 /Library/Application Support/IBM/SPSS/Statistics/19/CustomDialogs/<对话框名称> 文件夹中。

  如果为已安装对话框指定了备用位置（使用 SPSS_CDIALOGS_PATH 环境变量），则副本应位于相应位置的<对话框名称>文件夹下。

- ▶ 将副本重命名为<帮助文件>_<语言标识符>，请使用下表列出的语言标识。例如，如果帮助文件为 myhelp.htm，且您打算创建德语版本的文件，则本地化后的帮助文件应命名为 myhelp_de.htm。本地化帮助文件必须手动添加到您为对话框创建的定制对话框包装文件中。定制对话框包装文件为简单的 ZIP 文件，可以通过诸如 Windows 上的 WinZip 等应用程序打开并修改。

  如果存在其它需要本地化的辅助文件，例如图像文件，则需要手动修改主帮助文件中的相应路径，以指向本地化后的版本，这些辅助文件应与初始版本存放在一起。

当打开对话框时，SPSS Statistics 将搜索其语言标识符与当前语言匹配的帮助文件，当前语言在“选项”对话框的“常规”选项卡上的”语言”下拉列表中指定。如果找不到该帮助文件，则使用为对话框指定的帮助文件（在“对话框属性”的“帮助文件”属性中指定）。

语言标识符

de	德语
en	英语
es	西班牙语
fr	法语
it	意大利语
ja	日语
ko	朝鲜语
pl	波兰语
ru	俄语
zh_CN	简体中文
zh_TW	繁体中文

注意：定制对话框及其相关帮助文件中的文本并不限于 SPSS Statistics 所支持的语言。您可以采用任何语言写入对话框和帮助文本，无需创建特定于语言的属性和帮助文件。这样对话框的所有用户都可看到以该语言显示的文本。

章

20

生产作业

生产作业提供了以自动化方式运行 IBM® SPSS® Statistics 的功能。该程序可在无人看管的情况下运行，并在执行最后一条命令后终止，因此可以在其运行的同时执行其他任务或调度生产作业使其在预定的时间自动运行。如果常常运行相同的一组耗时的分析（例如周报告），则生产作业很有用。

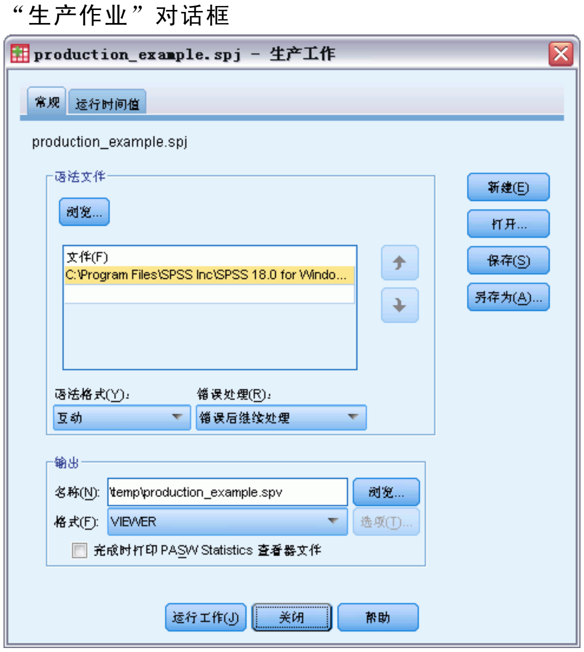
生产作业使用命令语法文件告诉 SPSS Statistics 该做什么。命令语法文件是包含命令语法的简单文本文件。可以使用任何文本编辑器创建该文件。您还可以通过将对话框选择粘贴到语法窗口中，或者通过编辑日志文件来生成命令语法。

创建生产作业：

▶ 从任意窗口中的菜单选择：

实用程序 > 生产工作

图片 20-1
“生产作业”对话框



▶ 单击**新建**以创建生产作业，或单击**打开**以打开现有生产作业。

注意：用低于 16.0 的版本创建的生产工具作业文件 (.spp) 在 16.0 或更高的版本中不能运行。现在提供了一个转换实用程序，可将 Windows 和 Macintosh 生产工具作业转换成生产作业 (.spj)。

- ▶ 选择一个或多个命令语法文件。
- ▶ 选择输出文件名称、位置和格式。
- ▶ 单击**保存**或**另存为**以保存生产作业。

语法格式。控制用于作业的语法规则的形式。

- ■ **交互式**。每条命令必须以句点结尾。句点可出现在命令中的任何位置，命令可延续多行，但作为某行最后一个非空格字符的句点被解释为命令结尾。接续行和新命令可在新行的任何位置开始。当您在语法窗口中选择和运行命令时，这些是生效的“交互式”规则。
- ■ **批处理**。每个命令必须在新的一行开始（在命令开头没有空格），接续行必须缩进至少一个空格。如果您要缩进新命令，可以使用加号、短划线或句点作为行开头处的第一个字符，然后再缩进实际命令。命令结尾处的句点是可选的。此设置与 INCLUDE 命令包括的命令文件的语法规则相兼容。

  注意：如果语法文件包含 GGRAPH 命令语法（包含 GPL 语句），请不要使用 Batch 选项。GPL 语句仅可在交互式规则下运行。

错误处理：控制作业中错误条件的处理。

- ■ **错误后继续处理**。作业中的错误不会自动停止命令处理。生产作业文件中的命令被当作正常命令流的一部分，命令处理以正常方式继续进行。
- ■ **立即停止处理**。当在生产作业文件中遇到第一个错误时，命令处理即停止。此设置与 INCLUDE 命令包括的命令文件的行为相兼容。

输出。控制生产作业结果的名称、位置和格式。可用格式选项有：

- ■ **查看器文件 (.spv)**。结果以 SPSS Statistics 查看器格式保存在指定的文件位置。您可以存储到磁盘或 IBM® SPSS® Collaboration and Deployment Services Repository。存储到 IBM SPSS Collaboration and Deployment Services Repository 需要 Statistics Adapter。
- ■ **Web 报告 (.spw)**。 结果存储到 IBM SPSS Collaboration and Deployment Services Repository。这需要 Statistics Adapter。
- ■ **Word/RTF (*.doc)**。枢轴表导出为 Word 表，并带有所有完整的格式属性，例如单元格边框、字体样式、背景色。文本输出导出为格式化的 RTF。图表、树形图和模型视图包括在 PNG 格式中。

  注意：Microsoft Word 可能无法正确显示过宽的表格。
- ■ **Excel (*.xls)**。枢轴表行、列和单元格导出为 Excel 行、列和单元格，并带有所有完整的格式属性，例如单元格边框、字体样式、背景色等等。文本输出连同所有字体属性一起导出。文本输出中的一行即为 Excel 文件中的一行，文本输出中行的所有内容包含在单个单元格中。图表、树形图和模型视图包括在 PNG 格式中。
- ■ **HTML (*.htm)**。枢轴表导出为 HTML 表。文本输出导出为预设置格式的 HTML。图表、树形图和模式视图按引用嵌入，您应以合适的格式导出图表，以包含在 HTML 文档中（例如，PNG 和 JPEG）。

- **可移植文件格式 (*.pdf)**。所有输出都将按“打印预览”中的显示导出，所有格式设置特性都不变。
- **PowerPoint 文件 (*.ppt)**。枢轴表导出为 Word 表，并嵌入到 PowerPoint 文件的各个幻灯片中，每个枢轴表对应一个幻灯片。将保留枢轴表的所有格式化属性，例如单元格边框、字体样式和背景色。图表、树形图和模型视图以 TIFF 格式导出。不包括文本输出。

  导出到 PowerPoint 仅在 Windows 操作系统上可用。
- **文本格式 (*.txt)**。文本输出格式包括纯文本、UTF-8 和 UTF-16。枢轴表可以以制表符分隔格式或空格分隔格式导出。所有文本输出以空格分隔格式导出。对于图表、树形图和模型视图，为每个图形将一行插入到文本文件中，以指示图像的文件名。

完成时打印查看器文件。在生产作业完成时，将最终查看器输出文件发送到打印机。

运行作业。这会在单独会话中运行生产作业。这在部署前检验新的生产作业时特别有用。

使用 OUTPUT 命令的生产作业

生产作业使用 OUTPUT 命令，例如 OUTPUT SAVE、OUTPUT ACTIVATE 和 OUTPUT NEW。生产作业期间执行的 OUTPUT SAVE 命令将把指定输出文档的内容写入到指定位置。这可作为生产作业创建的输出文件的补充。使用 OUTPUT NEW 创建新输出文档时，建议使用 OUTPUT SAVE 命令显式保存文档。

生产作业输出文件由作业结束时的活动输出文档的内容组成。对于包含 OUTPUT 命令的作业，输出文件可能不包含会话中创建的所有输出。例如，假设生产作业包括一些过程，后接 OUTPUT NEW 命令，然后是更多过程但没有 OUTPUT 命令。OUTPUT NEW 命令会定义一个新的活动输出文档。生产作业结束时，将仅包含 OUTPUT NEW 命令后执行的过程的输出。

HTML 选项

表格选项。HTML 格式无表格选项可用。所有枢轴表均转换为 HTML 表。

图像选项。可用图像类型有：EPS、JPEG、TIFF、PNG 和 BMP。在 Windows 操作系统上还可使用 EMF（增强型图元文件）格式。图像大小的缩放范围是从 1% 到 200%。

PowerPoint 选项

表格选项。可以使用查看器略图项作为幻灯片标题。每个幻灯片包含单个输出项。标题是从查看器概要窗格中的项的概要条目构成的。

图像选项。图像大小的缩放范围是从 1% 到 200%。（所有图像以 TIFF 格式导出到 PowerPoint。）

注意：PowerPoint 只在 Windows 操作系统上可用，且需要 PowerPoint 97 或更新版本。

PDF 选项

内嵌书签。此选项提供 PDF 文档中对应于浏览器概要条目的书签。与浏览器概要窗格一样，通过书签可以更轻松地浏览具有大量输出对象的文档。

内嵌字体。内嵌字体确保 PDF 文档在所有计算机上外观相同。否则，如果文档中使用的一些字体在用于查看（或打印）PDF 文档的计算机上不可用，则替换字体的效果可能欠佳。

文本选项

表格选项。枢轴表可以以制表符分隔格式或空格分隔格式导出。对于空格分隔格式，您还可以控制：

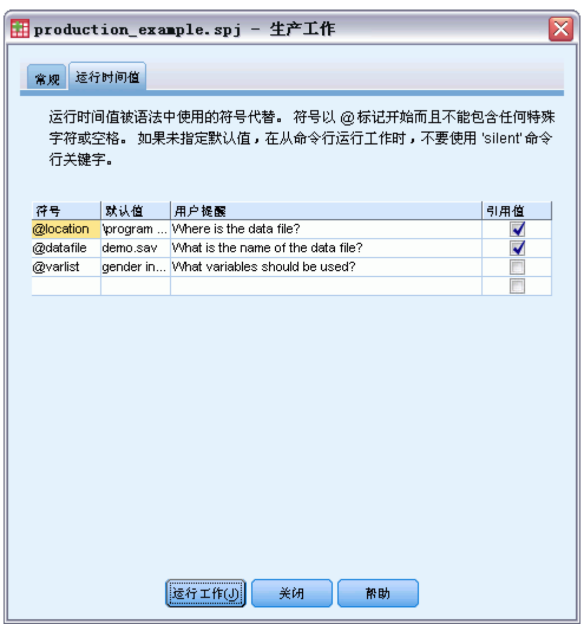
- **列宽**。**自动适应**不会对任何列内容换行，每列宽度为该列中最宽标签或值的宽度。**定制**为表格中所有列设置最大列宽，超过该宽度的值将折行显示在该列的下一行上。
- **行/列边框字符**。控制用于创建行和列边框的字符。如果不显示行和列边框，请将值留空。

图像选项。可用图像类型有：EPS、JPEG、TIFF、PNG 和 BMP。在 Windows 操作系统上还可使用 EMF（增强型图元文件）格式。图像大小的缩放范围是从 1% 到 200%。

运行时间值

在生产作业文件中定义、在命令语法文件中使用的运行时间值可简化一些任务，如对不同数据文件执行相同分析或对不同变量集运行相同的命令集等任务。例如，可以定义运行时间值@datafile以提示您，在每次运行使用字符串 @datafile 代替命令语法文件中的文件名的生产作业时，提供一个数据文件名。

图片 20-2
“生产作业”对话框，“运行时间值”选项卡



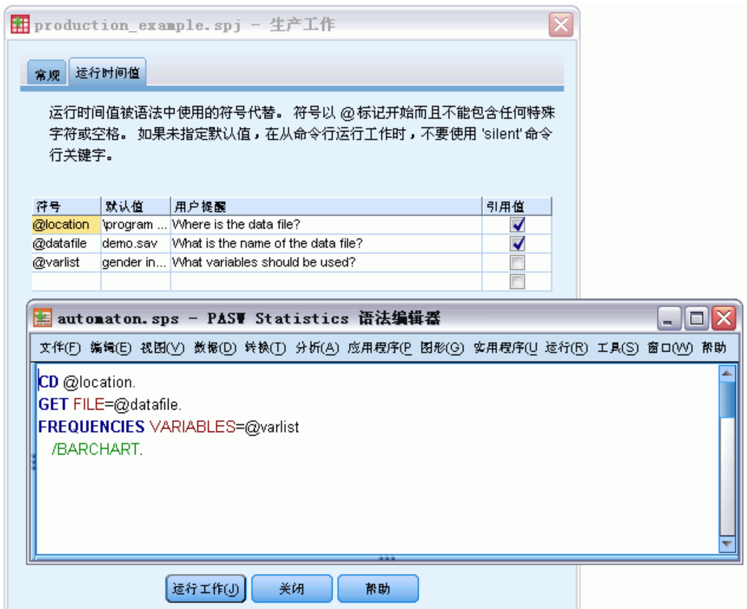
符号。在命令语法文件中使用的字符串，可触发生产作业提示用户输入值。所有符号名称必须以 @ 开头且必须符合变量命名规则。

默认值。在您不输入其他值的情况下，生产作业默认提供的值。当生产作业提示您输入信息时会显示该值。您可以在运行时替换或修改该值。如果未提供默认值，则在使用命令行开关运行生产作业时，不要使用 silent 关键字，除非另外使用了 -symbol 开关来指定运行时间值。

用户提示。在生产作业提示您输入信息时显示的说明性标签。例如，您可以使用短语“使用什么数据文件？”来标识一个需要填写数据文件名的字段。

引用值。使用引号将默认值或用户输入的值括起来。例如，文件指定必须以引号括起。

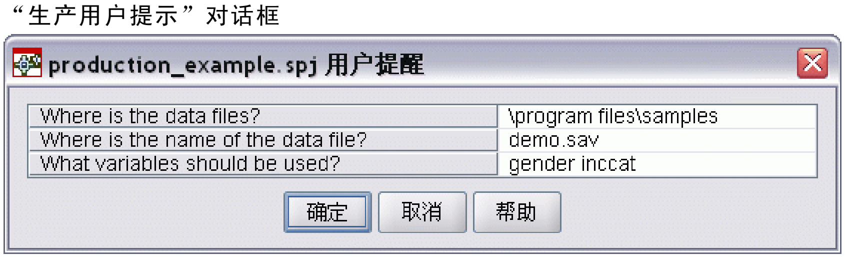
图片 20-3
命令语法文件中的运行时间符号



用户提示

只要您运行包含已定义运行时间符号的生产作业，生产作业就会提示您提供值。可以替换或修改显示的默认值。然后这些值将替换与生产作业相关联的所有命令语法文件中的运行时间符号。

图片 20-4
“生产用户提示”对话框



从命令行运行生产作业

命令行开关使您能够通过操作系统提供的一些调度工具来调度生产作业，使之在特定的时间自动运行。命令行参数的基本形式为：

```
stats filename.spj -production
```

根据生产作业的调用方式，可能需要包含 stats 可执行文件（位于应用程序的安装目录）的目录路径，和/或生产作业文件。

可以使用以下开关从命令行运行生产作业：

-production [prompt|silent]. 在生产模式下启动应用程序。prompt 和 silent 关键字指定是否显示对话框，以提示输入在作业中指定的运行时间值。prompt 关键字为缺省值，即显示对话框。silent 关键字不显示对话框。如果您使用 silent 关键字，则可使用 -symbol 开关定义运行时间符号。否则，将使用缺省值。在使用 -production 开关时，会忽略 -switchserver 和 -singleseat 开关。

-symbol <values>。在生产作业中使用的符号-值对列表。每个符号名均以 @ 开头。包含空格的值必须以引号括起。在字符串文本中包含引号或撇号的规则常常与操作系统有关，不过，使用双引号括起包含单引号或撇号的字符串的做法通常是可行的（例如，“'a quoted value'”）。

要在分布式分析模式下从远程服务器上运行生产作业，还需要指定服务器登录信息：

-server <inet:hostname:port>. 服务器的名称或 IP 地址以及端口号。仅 Windows。

-user <name>. 有效用户名。如果需要域名，则在用户名前面加上域名和反斜杠 (\)。仅 Windows。

-password <password>. 用户的密码。仅 Windows。

示例

```
stats \production_jobs\prodjob1.spj -production silent -symbol @datafile /data/July_data.sav
```

- 本例假设您从安装目录运行命令行，因此无需为 stats 可执行文件指定路径。
- 此外，它还假设生产作业指定了要引用的 @datafile 值（“运行时间值”选项卡中的“引用值”复选框），，因此在命令行中指定数据文件时无需使用引号。否则，您需要指定诸如 "'/data/July_data.sav'" 以使用引号括起数据文件指定，因为在命令语法中文件指定需要用引号括起。
- 生产作业位置的目录路径使用 Windows 反斜杠约定。在 Macintosh 和 Linux 上，则使用正斜杠。在引用数据文件指定中的正斜杠可适用于所有操作系统，因为该引用字符串会插入命令语法文件中。在所有操作系统上，对于包含文件指定的命令（例如，GET FILE、GET DATA、SAVE），正斜杠是可以接受的。
- silent 关键字会忽略生产文件中的所有用户提示。在生产作业包含的命令语法文件中，对于出现运行时间符号 @datafile 的位置，-symbol 开关会插入引用的数据文件名称和位置。

转换生产工具文件

用低于 16.0 的版本创建的生产工具作业文件 (.spp) 在 16.0 或更高的版本中不能运行。对于在之前版本中创建的 Windows 和 Macintosh 生产工具作业文件，可以使用安装目录下的 prodconvert 将其转换成新的生产作业文件 (.spj)。使用以下规格从命令窗口中运行 prodconvert：

```
[installpath]\prodconvert [filepath]\filename.spp
```

其中 [installpath] 是安装 IBM® SPSS® Statistics 的文件夹位置，[filepath] 是原始生产作业文件所在文件夹的位置。在与原始文件相同的文件夹下，将创建具有相同名称的新文件，但其扩展名为 .spj。（注意：如果路径包含空格，则用双引号将每个路径和文件规格括起来。在 Macintosh 操作系统上，使用正斜杠而非反斜杠。）

限制

- 不支持 WMF 和 EMF 图表格式。而使用 PNG 格式来替代这些格式。
- 不支持导出选项**输出文档**（**无图表**）、**仅图表**和**无**。包括选定格式支持的所有输出对象。
- 忽略远程服务器设置。要为分布式分析指定远程服务器设置，需要从命令行运行生产作业，并通过命令行开关指定服务器设置。
- 忽略“发布到 Web”设置。

章

21

输出管理系统

输出管理系统 (OMS) 能够用不同的格式将选中的输出类别自动写入不同的输出文件。格式包括：Word、Excel、PDF、IBM® SPSS® Statistics 数据文件格式 (.sav)、查看器文件格式 (.spv)、Web 报告格式 (.spw) 、XML、HTML 和文本。

使用输出管理系统控制面板

▶ 从菜单中选择：

实用程序 > OMS 控制面板...

图片 21-1
输出管理系统控制面板



可以使用该控制面板启动和停止输出到不同目标的路由。

- 每个 OMS 请求将保持活动状态，直到明确停止或直到会话结束为止。

- 在 OMS 请求中指定的目标文件对其他过程和其他应用程序不可用，直到该 OMS 请求终止为止。
- 如果一个 OMS 请求处于活动状态，则指定的目标文件存储在内存 (RAM) 中，因此将大量输出写入外部文件的活动 OMS 请求可能要消耗大量内存。
- 多个 OMS 请求是彼此独立的。相同的输出可以不同格式转到不同位置，这取决于不同 OMS 请求的指定。
- 任何特定目标中的输出对象的顺序都是创建这些对象的顺序，这是由生成输出的过程的顺序和操作确定的。

限制

- 对于输出 XML 格式，有关标题输出类型的指定不起任何作用。如果包含某个过程的任何输出，则应包含过程标题输出。
- 如果 OMS 指定导致对于过程只包含标题对象或注释表，则不会为此过程包含任何内容。

添加新 OMS 请求

▶ 选择要包括的输出类型（表、图形等等）。

▶ 选择要包括的命令。如果要包括所有输出，请选择列表中的所有项。

▶ 对于生成枢轴表输出的命令，选择要包括的特定表类型。

列表仅显示在选定命令中可用的表；在一个或多个选定命令中可用的任何表格类型都显示在列表中。如果没有选择任何命令，则显示所有表格类型。

▶ 要根据文本标签而不是子类型选择表，请单击**标签**。

▶ 单击**选项**指定输出格式（例如，SPSS Statistics 数据文件、XML 或 HTML）。默认情况下将使用输出 XML 格式。

▶ 指定输出目标：

- **文件**。所有选定的输出将转到单个文件。
- **根据对象名称**。输出将根据对象名称转到多个目标文件。将为每个输出对象分别创建一个文件，其文件名基于表子类型名称或表标签。输入目标文件夹名称。
- **新数据集**。对于 SPSS Statistics 数据文件格式输出，可以将输出转到数据集。可在相同的会话中继续使用数据集，但不会保存它，除非您在会话结束之前明确将其保存为文件。该选项仅适用于 SPSS Statistics 数据文件格式输出。数据集名称必须符合变量命名规则。

▶ 或者：

- 从查看器排除选定的输出。如果选择**排除在查看器之外**，则不会在“查看器”窗口中显示 OMS 请求的输出类型。如果多个活动的 OMS 请求包括相同的输出类型，则这些输出类型在查看器中的显示由包含这些输出类型的最近的 OMS 请求确定。
- 将 ID 字符串分配到请求。所有请求都自动分配有一个 ID 值，您可以使用描述性 ID 覆盖系统默认的 ID 字符串，如果您有多个活动请求并且想轻松标识它们，那么此方法很有用。分配的 ID 值不能以美元符号 ($) 开头。

下列提示适用于选择列表中的多个项：

- 按 Ctrl 和 A 可选择列表中的所有项。
- 按住 Shift 并单击可选择多个连续的项。
- 按住 Ctrl 并单击可选择多个不连续的项。

终止和删除 OMS 请求

“请求”列表中显示活动的和新的 OMS 请求，最新的请求显示在顶部。您可以通过单击并拖动边界来更改信息列的宽度，可以水平滚动列表来查看有关特定请求的更多信息。

在状态列的词 Active 之后的星号 (*) 表示使用命令语法创建的、包含“控制面板”中不可用的功能的 OMS 请求。

终止特定的活动 OMS 请求：

▶ 在“请求”列表中，单击请求的行中的任意单元格。

▶ 单击**终止**。

终止所有的活动 OMS 请求：

▶ 单击**全部终止**。

删除新请求（已添加但尚未成为活动状态的请求）：

▶ 在“请求”列表中，单击请求的行中的任意单元格。

▶ 单击**删除**。

注意：单击**确定**后，活动 OMS 请求才会终止。

输出对象类型

有不同类型的输出对象：

图表。 这包括使用图表生成器创建的图表、图表过程以及通过统计过程创建的图表（例如，通过频率过程创建的条形图）。

标题。在查看器的概要窗格中标记为标题的文本对象。

日志。日志文本对象。日志对象包含特定类型的错误和警告消息。根据“选项”设置（“编辑”菜单，“选项”，“查看器”选项卡），日志对象还可以包含在会话期间执行的命令语法。日志对象在查看器的概要窗格中标记为日志。

模型。 模型查看器中显示的输出对象。单个模型对象可以包含多个模型视图，包括表格和图表。

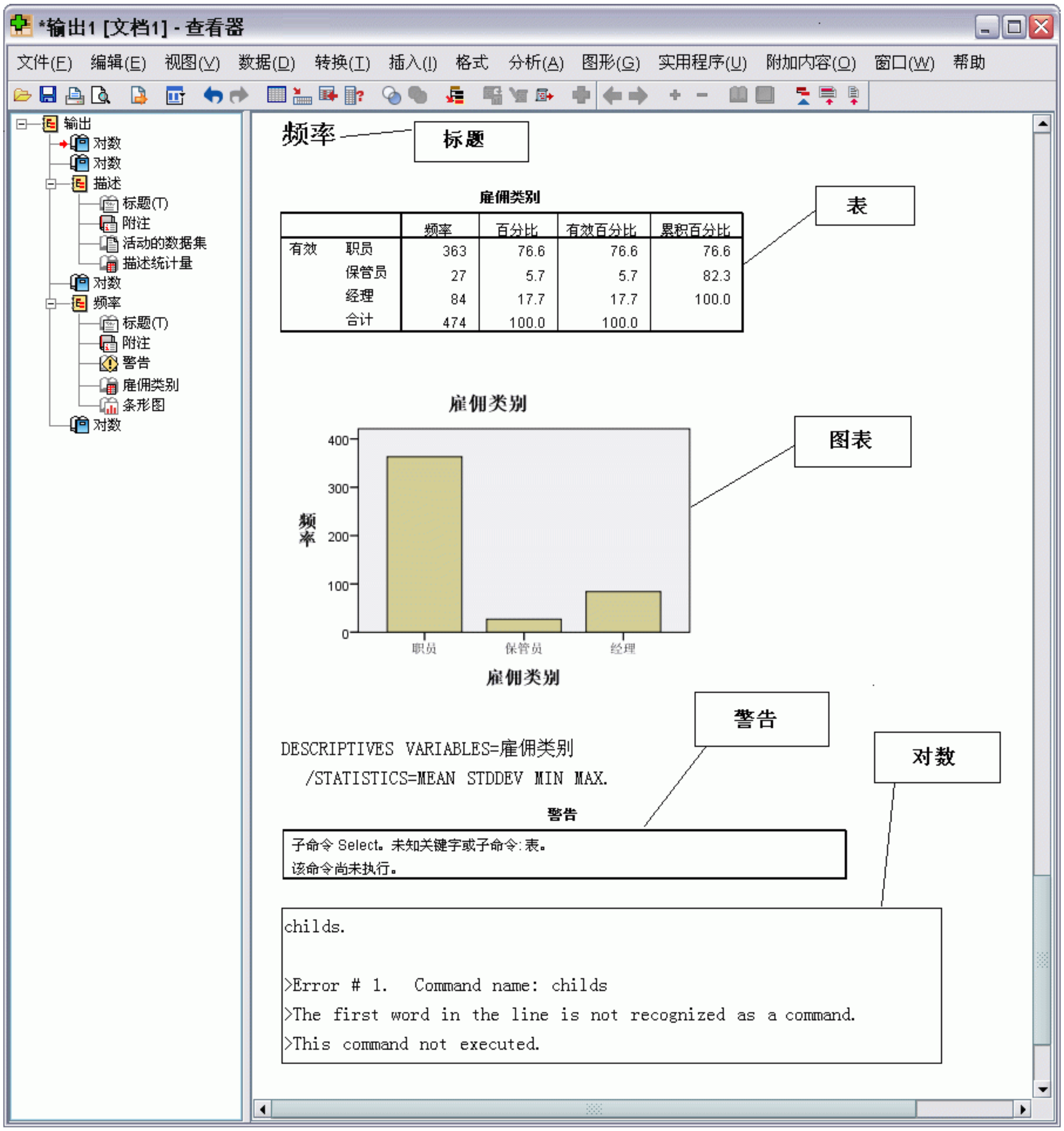
表。在查看器中作为枢轴表的输出对象（包括“注释”表）。表是可以转到 IBM® SPSS® Statistics 数据文件 (.sav) 格式的唯一输出对象。

文本。不是日志也不是标题的文本对象（包括在查看器的概要窗格中标记为文本输出的对象）。

树。由“决策树”选项生成的树模型图。

警告。警告对象包含特定类型的错误和警告消息。

图片 21-2
输出对象类型



命令标识符和表子类型

命令标识符

命令标识符可用于所有统计和绘图过程，以及在查看器的概要窗格中生成包含自己的可识别标题的输出区的任何其他命令。这些标识符通常（但不总是）与菜单上的过程名称和对话框标题相同或类似，这些名称和标题通常（但不总是）与基础命令名称相同。例如，“频率”过程的命令标识符为“Frequencies”，基础命令名称也相同。

但是，有时候过程名称和命令标识符和/或命令名称并不相同。例如，“非参数检验”子菜单（位于“分析”菜单中）上的所有过程使用相同的基础命令，命令标识符与基础命令名称“Npar 检验”相同。

表子类型

表子类型是可以生成的枢轴表的不同类型。某些子类型仅由一个命令生成；其他子类型可由多个命令生成（尽管表看上去并不相同）。虽然表子类型名称通常是描述性的，但也可以选择其他名称（尤其是选择了大量命令时）；并且，两个子类型的名称可以非常类似。

查找命令标识符和表子类型

如果不确定，可以在“查看器”窗口中查找命令标识符和表子类型名称：

▶ 运行过程以在查看器中生成一些输出。

▶ 右键单击查看器概要窗格中的项。

▶ 选择**复制 OMS 命令标识符**或**复制 OMS 表子类型**。

▶ 将复制的命令标识符或表子类型名称粘贴到任意文本编辑器中（如语法编辑器窗口）。

标签

作为表子类型名称的替代选项，您可以根据查看器概要窗格中显示的文本选择表。还可以根据其他对象类型的标签来选择它们。标签可以用于区分同一类型的多个表，在这些表中，概要文本反映特定输出对象（如变量名或标签）的某个属性。但是有一些因素会影响标签文本：

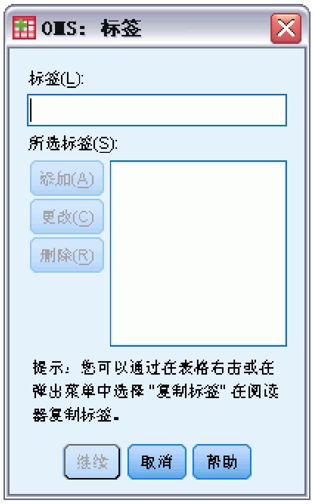
- 如果正在处理拆分文件，则拆分文件组标识将会附加到标签上。
- 包含有关变量或值的信息的标签受当前输出标签选项设置（“编辑”菜单，“选项”，“输出标签”选项卡）的影响。
- 标签受当前输出语言设置（“编辑”菜单，“选项”，“常规”选项卡）的影响。

指定用于标识输出对象的标签

▶ 在输出管理系统控制面板中，选择一个或多个输出类型，然后选择一个或多个命令。

▶ 单击**标签**。

图片 21-3
“OMS：标签”对话框



▶ 输入与显示在“查看器”窗口的概要窗格中的标签完全相同的标签。（也可以右键单击概要中的项，选择**复制 OMS 标签**，然后将复制的标签粘贴到“标签”文本字段中。）

▶ 单击**添加**。

▶ 对要包括的每个标签重复该过程。

▶ 单击**继续**。

通配符

可以使用星号 (*) 作为通配符字符，用作标签字符串的最后一个字符。将选择以指定字符串（星号除外）开头的所有标签。此过程只在星号是最后一个字符时才起作用，因为星号可显示为标签中的有效字符。

OMS：选项

您可以使用“OMS：选项”对话框来：

■ 指定输出格式。

■ 指定图像格式（对于 HTML 和输出 XML 输出格式）。

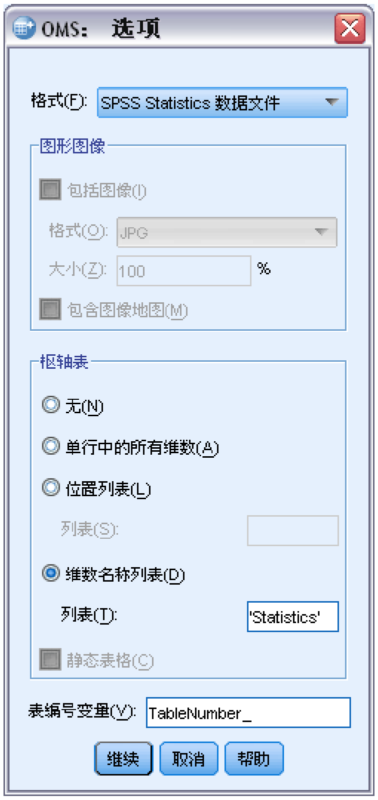
■ 指定哪些表维度元素应转到行维度中。

■ 对于 IBM® SPSS® Statistics 数据文件格式，将包括标识作为每个个案的源的表序列号的变量。

指定 OMS 选项

▶ 单击输出管理系统控制窗格中的**选项**。

图片 21-4
“OMS：选项”对话框



格式

Excel。 Excel 97-2003 格式。 枢轴表行、列和单元格导出为 Excel 行、列和单元格，并带有所有完整的格式属性，例如单元格边框、字体样式、背景色等等。文本输出连同所有字体属性一起导出。文本输出中的一行即为 Excel 文件中的一行，文本输出中行的所有内容包含在单个单元格中。图表、树形图和模型视图包括在 PNG 格式中。

HTML。将在查看器中作为枢轴表的输出对象被转换为简单的 HTML 表。不支持任何表格外观属性（字体特征、边框样式、颜色等等）。文本输出对象在 HTML 中带有标记 <PRE>。图表、树形图和模型视图将以选定的图形格式导出为单独的文件，并通过引用嵌入。图像文件名称将 HTML 文件名作为根名称，其后用以 0 开头的连续整数。

输出 XML。符合 spss 输出架构的 XML。

PDF。 输出将按“打印预览”中的显示导出，所有格式设置特性都不变。PDF 文件包含对应于查看器概要窗格中条目的书签。

SPSS Statistics 数据文件。此格式是二进制文件格式。将排除表以外的所有输出对象类型。表的每列成为数据文件中的一个变量。要使用在相同会话中通过 OMS 创建的数据文件，必须先终止活动的 OMS 请求，然后才能打开数据文件。

文本。空格分隔的文本。 输出写为文本，具有与固定间距字体的间距对齐的表格输出。 排除图表、树形图和模型视图。

Tab 文本。制表符分隔的文本。 对于在查看器中显示为枢轴表的输出，制表符分隔表列元素。文本区线按原样写出；不尝试在有用的位置使用制表符分隔它们。 排除图表、树形图和模型视图。

查看器文件。 这与保存“查看器”窗口内容时使用的格式相同。

Web 报告文件。 该输出文件格式设计与 Predictive Enterprise Services 配合使用。其本质上与 SPSS Statistics 查看器格式相同，不同之处在于树形图以静态图像形式存储。

Word/RTF。 枢轴表导出为 Word 表，并带有所有完整的格式属性，例如单元格边框、字体样式、背景色。文本输出导出为格式化的 RTF。图表、树形图和模型视图包括在 PNG 格式中。

图形图像

对于 HTML 和输出 XML 格式，可以包括图表、树形图和模型视图作为图像文件。为每个图表和/或树创建单独的图像文件。

- 对于 HTML 文档格式，在 HTML 文档中为每个图像文件包含标准的 <IMG SRC='filename'> 标记。
- 对于输出 XML 文档格式，XML 文件为每个图像文件包含一个**图表**元素并带有 ImageFile 属性，一般格式为 <chart imageFile="filepath/filename"/>。
- 图像文件保存在单独的子目录（文件夹）中。子目录的名称是目标文件的名称，没有任何扩展名，在末尾附加有 _files。例如，如果目标文件是 julydata.htm，则图像子目录的名称将是 julydata_files。

格式。可用的图像格式有 PNG、JPG、EMF、BMP 和 VML。

- EMF（增强型图元文件）格式只在 Windows 操作系统上可用。
- VML 图像格式只对 HTML 文档格式可用。
- VML 图像格式不创建单独的图像文件。图像着色的 VML 代码嵌入在 HTML 中。
- VML 图像格式不包括树形图。

大小。图像大小的缩放范围是从 10% 到 200%。

包含图像地图。对于 HTML 文档格式，该选项可创建图像地图工具提示，它显示某些图表元素例如线图上选定点或条形图上选定条的值信息。

枢轴表

对于枢轴表输出，可以指定应显示在列中的维度元素。所有其他维度元素显示在行中。对于 SPSS Statistics 数据文件格式，表列成为变量，行成为个案。

- 如果指定列的多个维度元素，它们将以列出的顺序嵌套在列中。对于 SPSS Statistics 数据文件格式，变量名称由内嵌列元素构成。
- 如果表不包含任何列出的维度元素，则该表的所有维度元素将显示在行中。
- 在此处指定的表旋转对显示在查看器中的表没有影响。

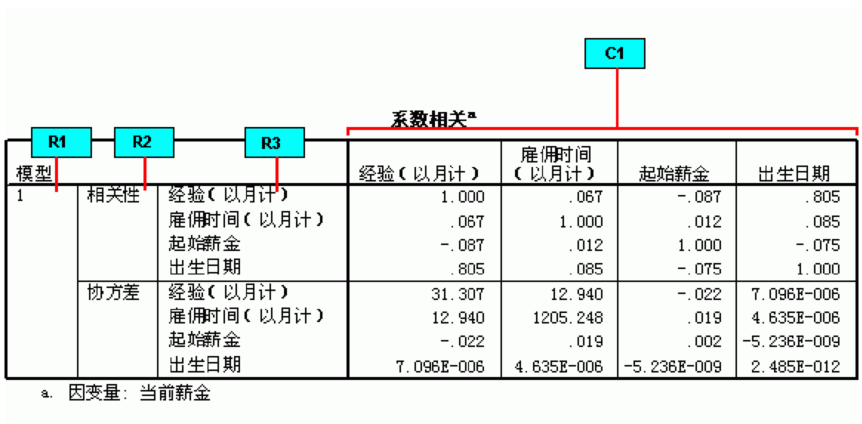
表的每个维度 — 行、列、层 — 可包含零个元素或多个元素。例如，简单二维交叉制表包含一个行维度元素和一个列维度元素，每个元素包含表中使用的一个变量。您可以使用位置参数或维度元素“名称”来指定要放在列维度中的维度元素。

行中的所有维度。为每个表创建一行。对于 SPSS Statistics 格式数据文件，这意味着每个表是一个个案，所有表元素都是变量。

位置列表。位置参数的一般形式是一个字母，表示元素的缺省位置，即 C 表示列，R 表示行，L 表示层；后面是一个正整数，表示该维度中的默认位置。例如，R1 表示最外面的行维度元素。

- 要指定多个维度中的多个元素，请使用空格分隔每个维度，例如 R1 C2。
- 跟有 ALL 的维度字母表示该维度中按照默认顺序排列的所有元素。例如，CALL 与缺省行为相同（使用按照默认顺序排列的所有列元素来创建列）。
- CALL RALL LALL（或 RALL CALL LALL 等等）将把所有维度元素放在列中。对于 SPSS Statistics 数据文件格式，此操作将在数据文件的每个表中创建一行/个案。

图片 21-5
行位置参数和列位置参数



系数相关[a]

模型			经验(以月计)	雇佣时间(以月计)	起始薪金	出生日期
1	相关性	经验(以月计)	1.000	.067	-.087	.805
		雇佣时间(以月计)	.067	1.000	.012	.085
		起始薪金	-.087	.012	1.000	-.075
		出生日期	.805	.085	-.075	1.000
	协方差	经验(以月计)	31.307	12.940	-.022	7.096E-006
		雇佣时间(以月计)	12.940	1205.248	.019	4.635E-006
		起始薪金	-.022	.019	.002	-5.236E-009
		出生日期	7.096E-006	4.635E-006	-5.236E-009	2.485E-012

a. 因变量：当前薪金

维度名称列表。作为位置参数的替代选择，可以使用维度元素“名称”，这些名称是显示在表中的文本标签。例如，简单二维交叉制表包含单个行维度元素和单个列维度元素，每个元素都具有基于维度中的变量的标签，加上标记为 Statistics 的单个层维度元素（如果英语是输出语言）。

- 维度元素名称可能不同，这取决于影响表中变量名和/或标签的显示的输出语言和/或设置。
- 每个维度元素名称必须用单引号或双引号括起来。要指定多个维度元素名称，请在每个用引号括起的名称之间加一个空格。

与维度元素关联的标签不总是明显的。

查看枢轴表的所有维度元素及其标签

▶ 激活（双击）查看器中的表。

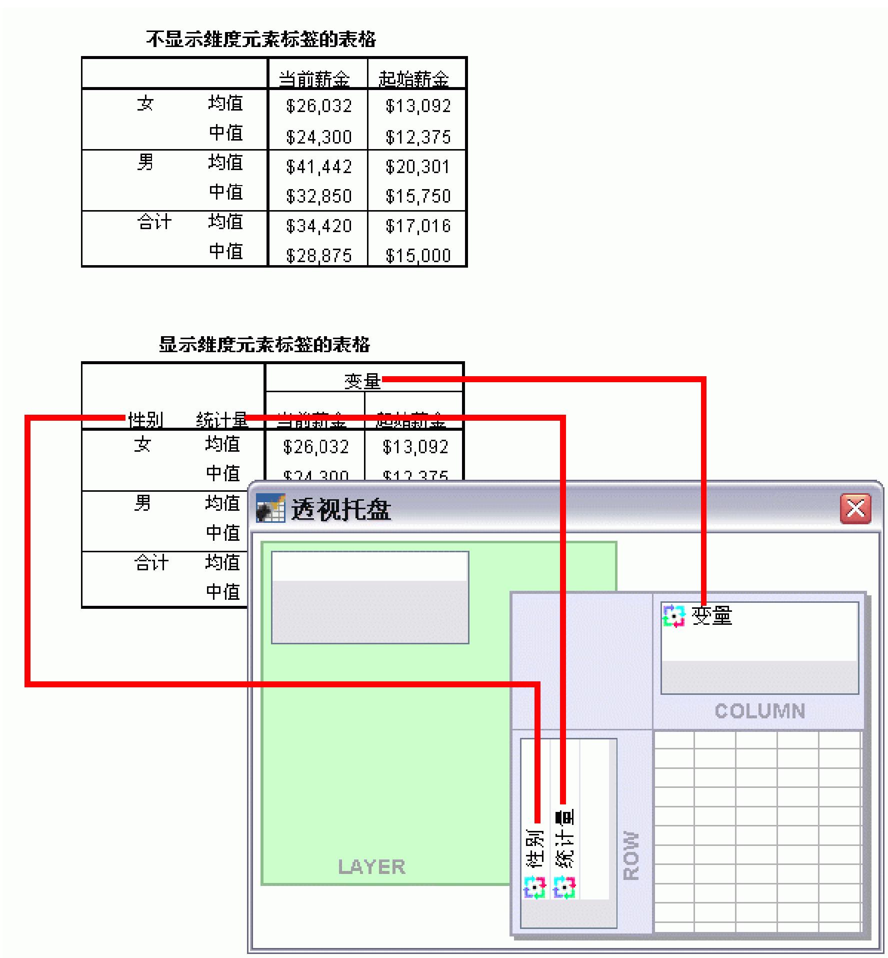
▶ 从菜单中选择：
视图 > 全部显示

和/或

▶ 如果未显示旋转托盘，请从菜单中选择：
透视 > 透视托盘

元素标签显示在旋转托盘中。

图片 21-6
显示在标签和旋转托盘中的维度元素名称



日志记录

您可以将 OMS 活动以 XML 或文本格式记录在日志中。

- 该日志将跟踪会话的所有新 OMS 请求，但不包括在请求记录前已激活的 OMS 请求。
- 如果指定新的日志文件或者取消选择（取消选中）**记录 OMS 活动**，则当前日志文件将结束。

指定 OMS 日志记录：

▶ 单击输出管理系统控制窗格中的**日志记录**。

从查看器排除输出显示

排除在查看器之外复选框影响在 OMS 请求中选择的所有输出，禁止在查看器窗口中显示这些输出。对于生成大量输出的生产作业，以及当您不需要查看器文档形式（.spv 文件）的结果时，此过程通常很有用。还可以使用此功能取消显示不想看到的特定输出对象，而不会将任何其他输出转到某些外部文件和格式。

取消显示某些输出对象而不将其他输出转到外部文件：

▶ 创建标识不需要的输出的 OMS 请求。

▶ 选择**排除在查看器之外**。

▶ 对于输出目标，选择**文件**，但将“文件”字段留空。

▶ 单击**添加**。

选定的输出将从查看器中排除，而所有其他输出将以正常形式显示在查看器中。

注意：该设置对保存到外部格式或文件（包括查看器 SPV 和 SPW 格式）的 OMS 输出没有影响。它对在通过批处理工具（随 IBM® SPSS® Statistics 服务器提供）执行的批处理作业中保存到 SPV 格式的输出也没有影响。

将输出转到 IBM SPSS Statistics 数据文件

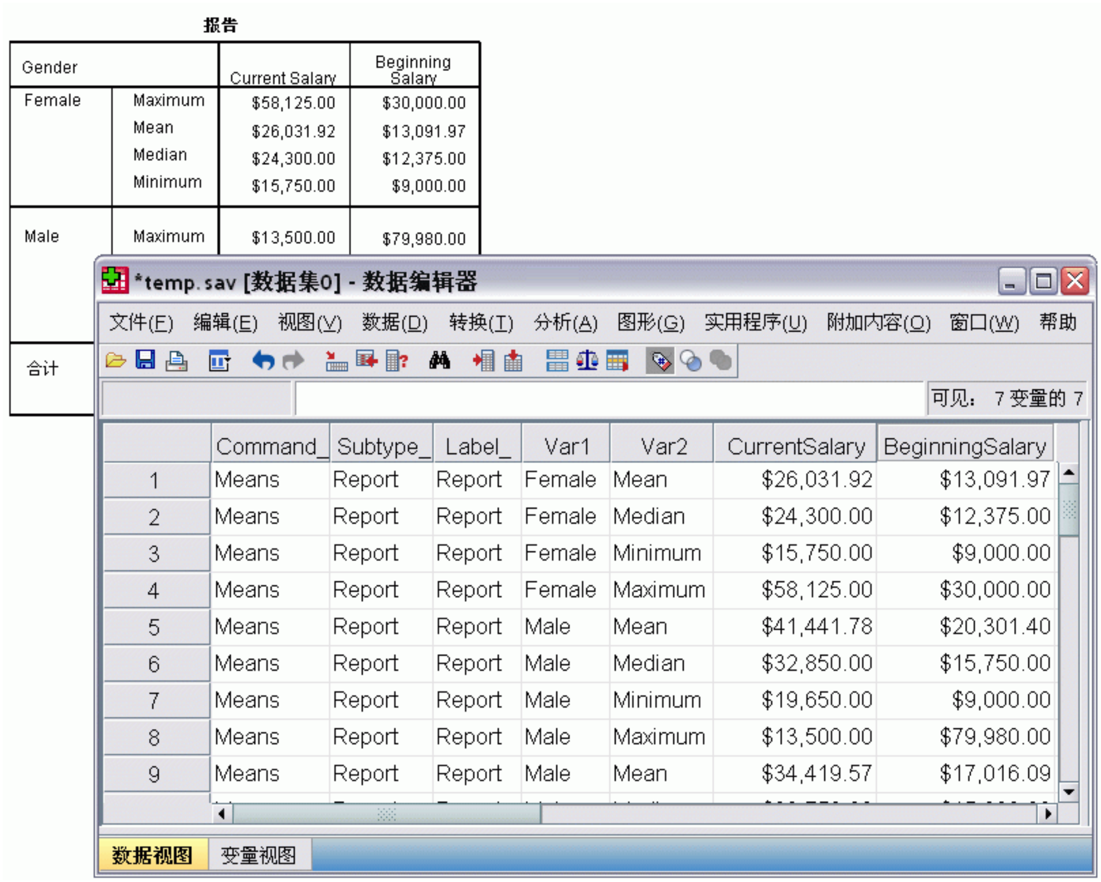
IBM® SPSS® Statistics 格式的数据文件由列中的变量和行中的个案组成，这是枢轴表转换为数据文件所用的基本格式：

- 表中的列是数据文件中的变量。有效变量名从列标签构成。
- 表中的行标签成为数据文件中具有一般变量名的变量（Var1、Var2、Var3，等等）。这些变量的值是表中的行标签。
- 数据文件中将自动包括三个表标识符变量：Command_、Subtype_ 和 Label_。所有这三个变量都是字符串变量。前两个变量与命令和子类型标识符对应。 Label_ 包含表标题文本。
- 表中的行成为数据文件中的个案。

示例:单个二维表

在最简单的个案（单个二维表）中，表列成为变量，行成为数据文件中的个案。

图片 21-7
单个二维表

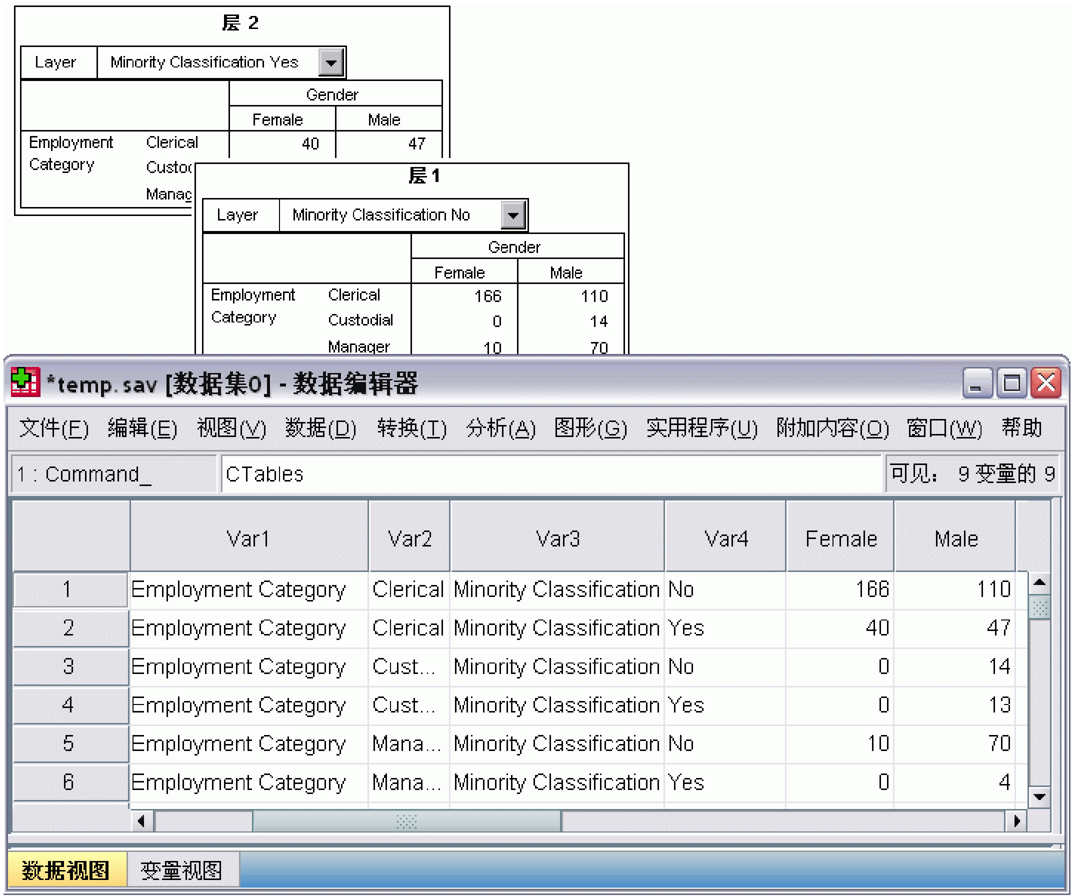


- 前三个变量根据命令、子类型和标签标识源表。
- 定义表中的行的两个元素（变量性别和统计度量的值）将分配一般变量名 Var1 和 Var2。这些变量都是字符串变量。
- 表中的列标签用于创建有效的变量名。在这种情况下，这些变量名都基于表中汇总的三个刻度变量的变量标签。如果变量没有已定义的变量标签，或者您选择显示变量名而不是变量标签作为表中的列标签，则新数据文件中的变量名应与源数据文件中的变量名相同。

示例:带有层的表

除了行和列以外，表还可以包含第三个维度：层维度。

图片 21-8
带有层的表



- 在表中，标记为少数分类的变量定义层。在数据文件中，将创建两个附加变量：一个变量标识层元素，另一个变量标识层元素的类别。
- 与从行元素创建的变量一样，从层元素创建的变量也是字符串变量，具有一般变量名（前缀 Var 后面是序列号）。

从多个表创建的数据文件

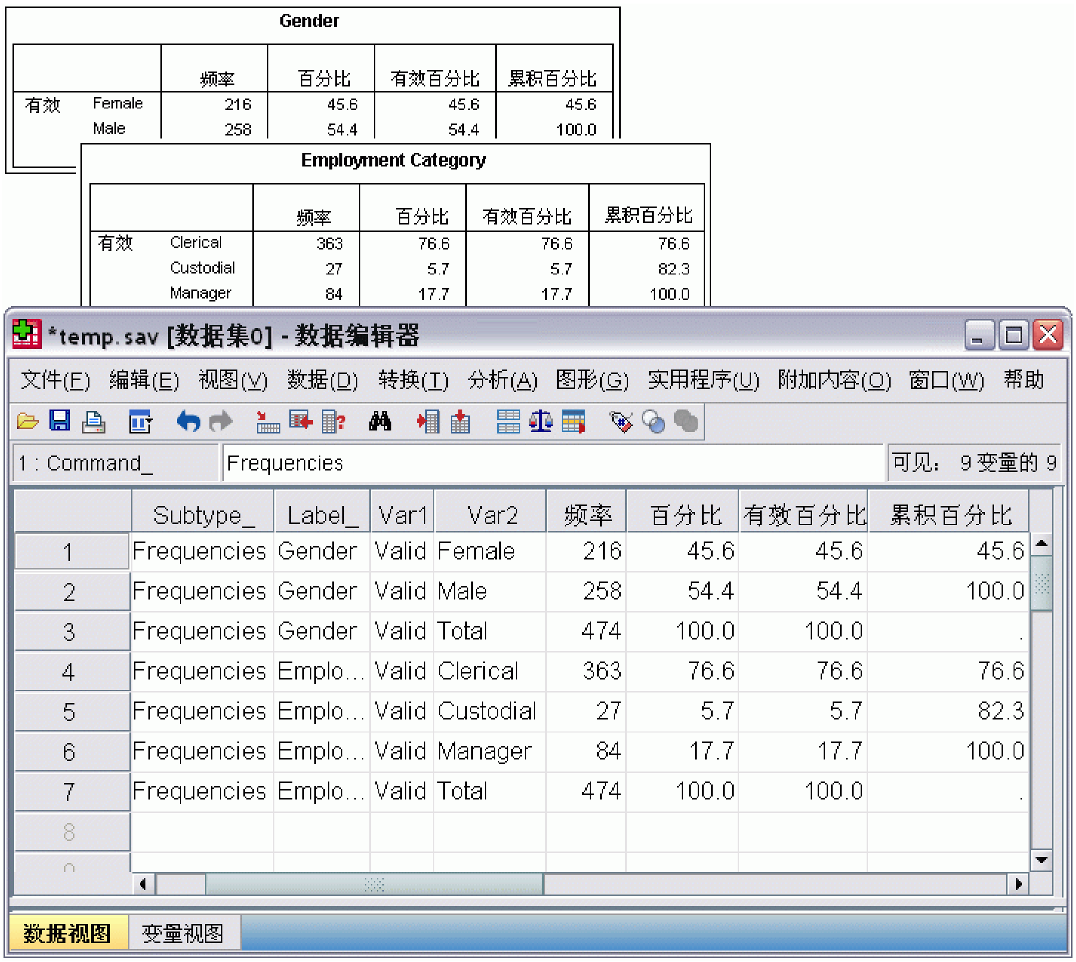
将多个表转到相同的数据文件时，每个表以类似于合并数据文件的方式添加到数据文件，方法是将一个数据文件的个案添加到另一个数据文件（“数据”菜单，“合并文件”，“添加个案”）。

- 每个后续的表将始终向数据文件添加个案。
- 如果表中的列标签不同，每个表也可以将变量添加到数据文件，对于没有相同列标签的其他表包含个案缺失值。

示例:具有相同列标签的多个表

包含相同列标签的多个表通常生成可立即使用的数据文件（不需要其他操作的数据文件）。例如，“频率”过程的两个或更多频率表将具有相同的列标签。

图片 21-9
具有相同列标签的两个表



Gender

		频率	百分比	有效百分比	累积百分比
有效	Female	216	45.6	45.6	45.6
	Male	258	54.4	54.4	100.0

Employment Category

		频率	百分比	有效百分比	累积百分比
有效	Clerical	363	76.6	76.6	76.6
	Custodial	27	5.7	5.7	82.3
	Manager	84	17.7	17.7	100.0

*temp.sav [数据集0] - 数据编辑器

文件(F) 编辑(E) 视图(V) 数据(D) 转换(T) 分析(A) 图形(G) 实用程序(U) 附加内容(O) 窗口(W) 帮助

1 : Command_ Frequencies 可见: 9 变量的 9

	Subtype_	Label_	Var1	Var2	频率	百分比	有效百分比	累积百分比
1	Frequencies	Gender	Valid	Female	216	45.6	45.6	45.6
2	Frequencies	Gender	Valid	Male	258	54.4	54.4	100.0
3	Frequencies	Gender	Valid	Total	474	100.0	100.0	.
4	Frequencies	Emplo...	Valid	Clerical	363	76.6	76.6	76.6
5	Frequencies	Emplo...	Valid	Custodial	27	5.7	5.7	82.3
6	Frequencies	Emplo...	Valid	Manager	84	17.7	17.7	100.0
7	Frequencies	Emplo...	Valid	Total	474	100.0	100.0	.
8								

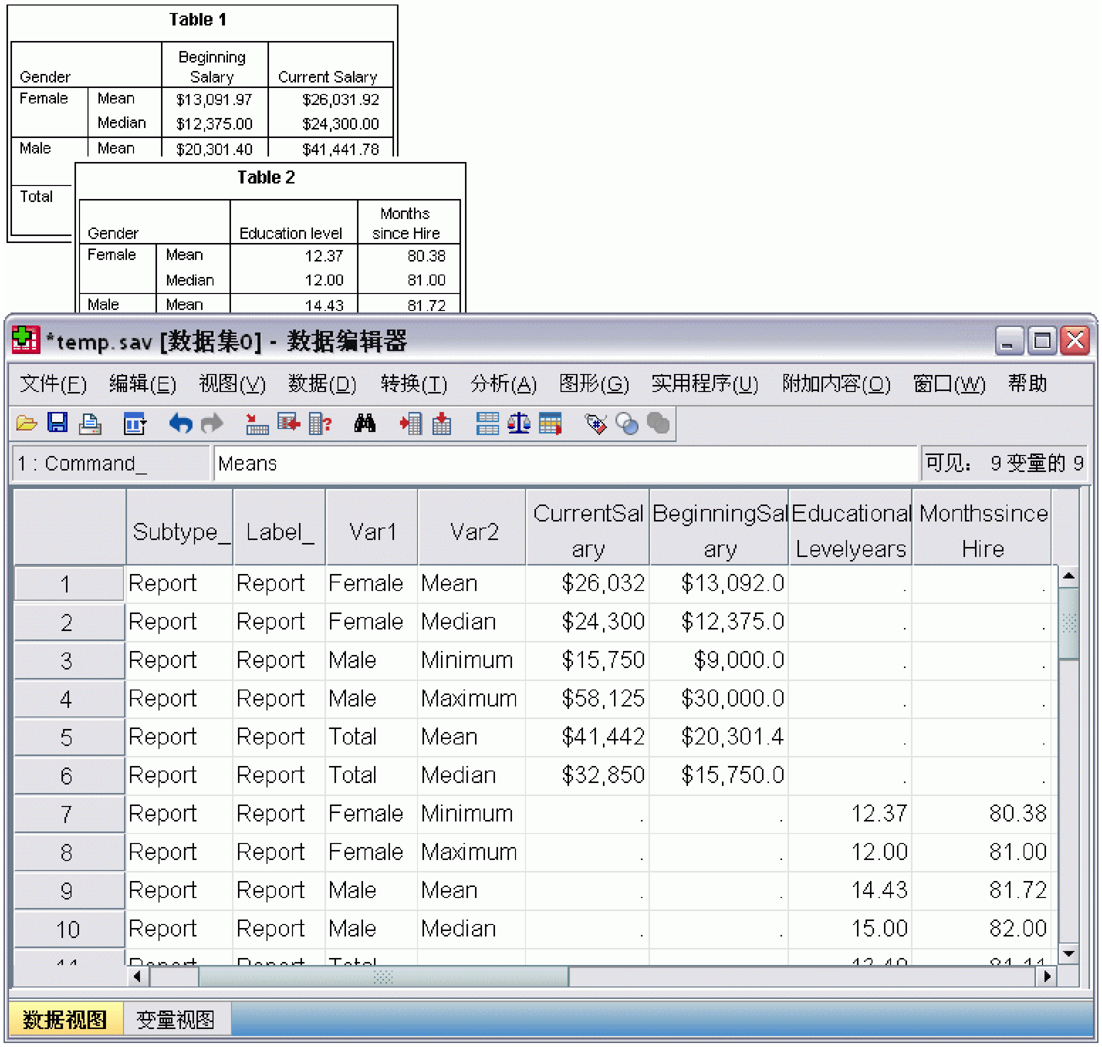
数据视图 变量视图

- 第二个表将附加个案（行）添加到数据文件，但不添加新变量，因为列标签是完全相同的，因此没有缺失数据区。
- 虽然 Command_ 和 Subtype_ 的值是相同的，但是 Label_ 值标识每组个案的源表，因为两个频率表的标题不同。

示例:具有不同列标签的多个表

将在数据文件中为表中每个唯一的列标签创建一个新变量，这些表将转到数据文件。如果表包含不同的列标签，则此过程将导致出现缺失值区。

图片 21-10
具有不同列标签的两个表



- 第一个表中的列标记为起始薪金和当前薪金，这在第二个表中没有，因此将导致第二个表中这些个案的变量的缺失值。
- 相反，第二个表中的列标记为教育水平和雇用时间，这在第一个表中没有，因此将导致第一个表中这些个案的变量的缺失值。
- 即使表的子类型相同，也会存在不匹配的变量，就像此示例中的变量。在此示例中，这两个表的子类型都相同。

示例：未从多个表创建的数据文件

如果表的行元素的数量互不相同，则不会创建任何数据文件。行数不必相同；成为数据文件中的变量的行元素的数量必须相同。例如，两个变量的交叉制表和三个变量的交叉制表包含不同数量的行元素，因为“层”变量实际上嵌套到默认的三个变量交叉制表显示的行变量中。

图片 21-11
具有不同数量的行元素的表

雇佣类别 * 性别 交叉制表

		性别		合计
		女	男	
雇佣类别	职员	206	157	363
	保管员	0	27	27
	经理	10	74	84
合计		216	258	474

雇佣类别 * 性别 * 少数民族分类 交叉制表

少数民族分类			性别		合计
			女	男	
否	雇佣类别	职员	166	110	276
		保管员	0	14	14
		经理	10	70	80
	合计		176	194	370
是	雇佣类别	职员	40	47	87
		保管员	0	13	13
		经理	0	4	4
	合计		40	64	104

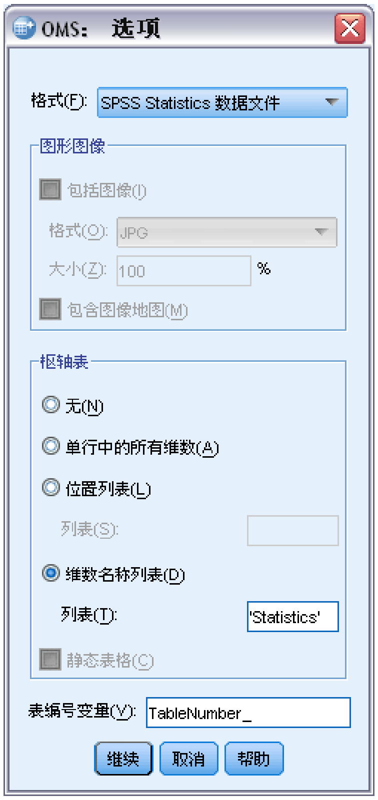
控制列元素转换为数据文件中的控制变量

在输出管理系统控制面板的“选项”对话框中，可以指定列中应包括的维度元素，以便用于创建生成的数据文件中的变量。此过程等同于在查看器中旋转表。

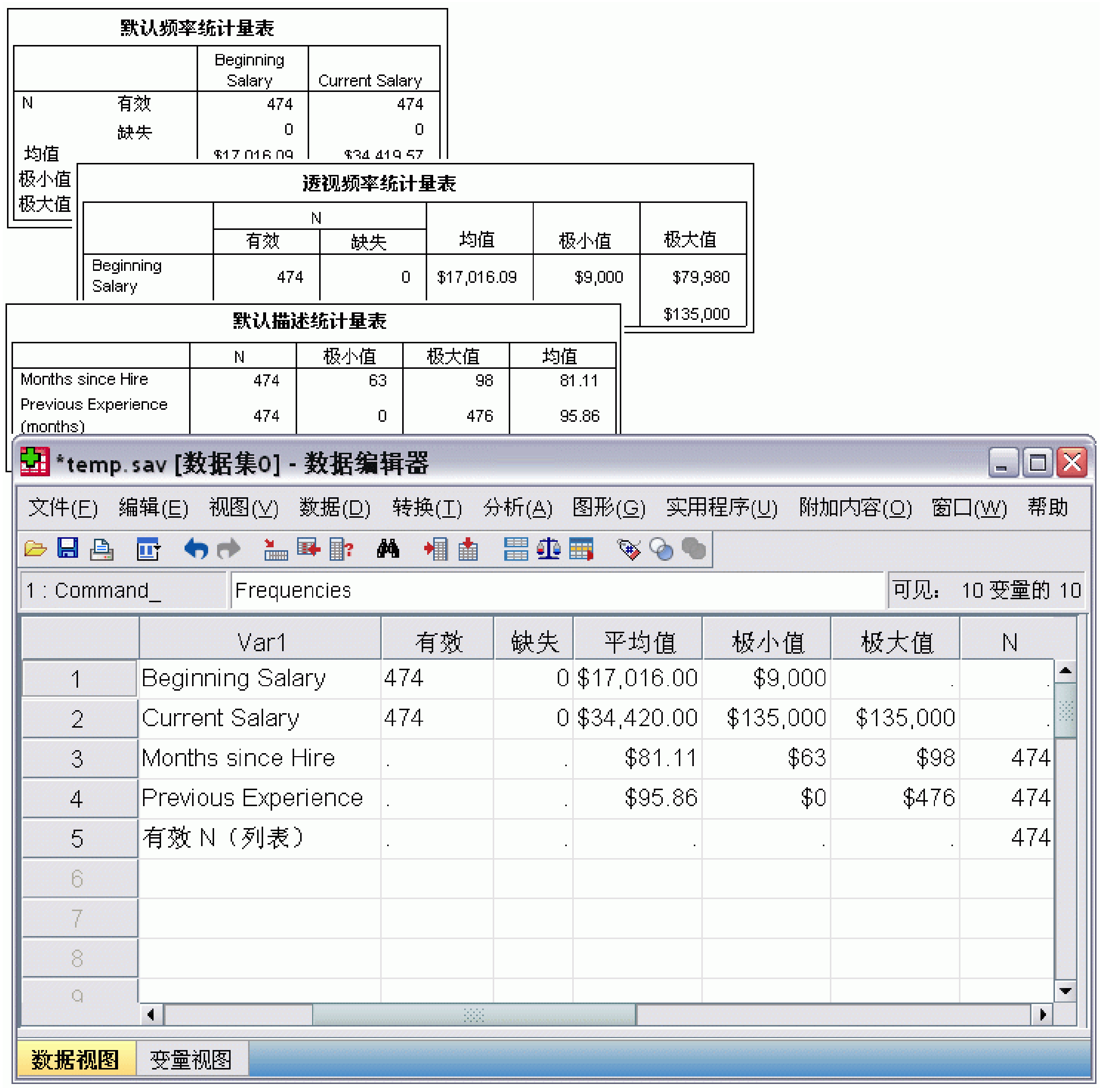
例如，“频率”过程生成一个描述性统计表，其中统计量在行中，而“描述统计”过程生成描述性统计表，其中统计量在列中。要以有意义的方式同时包括相同数据文件中的这两个表类型，您需要更改一种表类型的列维度。

因为这两个表类型使用元素名称“统计量”作为统计维度，所以可将“频率”统计表中的统计量放在列中，方法是在“OMS：选项”对话框的维度名称列表中指定“统计量”（在引号中）。

图片 21-12
“OMS：选项”对话框



图片 21-13
通过旋转维度元素来组合数据文件中不同的表类型



某些变量将具有缺失值，因为表结构仍与列中的统计量不完全相同。

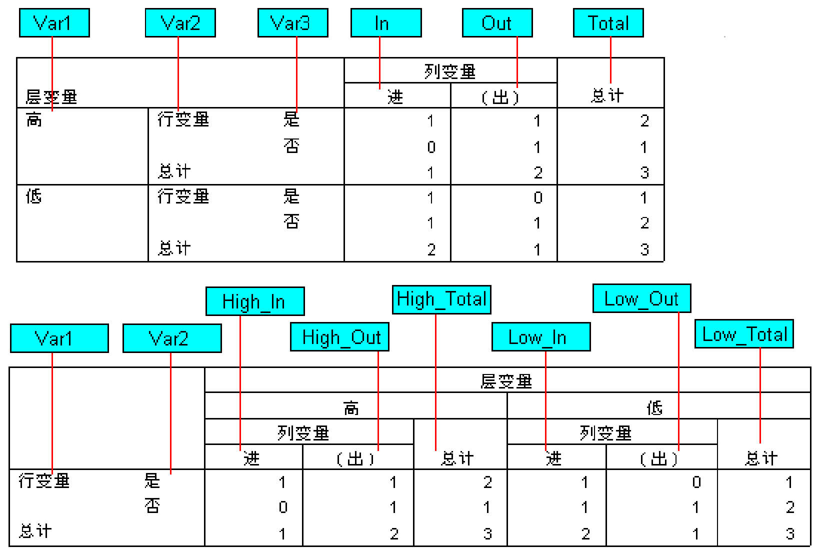
OMS 生成的数据文件中的变量名

OMS 从列标签构建有效的唯一变量名：

- 行元素和层元素分配有一般变量名：前缀 Var 后面是一个序列号。
- 将移去变量名中不允许的字符（空格、括号等等）。例如，“This (Column) Label”将成为名为 ThisColumnLabel 的变量。
- 如果标签以变量名中允许的但不允许作为第一个字符的字符开头（例如数字），则将插入“@”作为前缀。例如，“2nd”将成为名为 @2nd 的变量。

- 将从生成的变量名中移去标签末尾的下划线或句点。不移去位于自动生成的变量 Command_、Subtype_ 和 Label_ 末尾的下划线。
- 如果列维度中有多个元素，则可通过在类别标签之间添加下划线来组合类别标签，以此构建变量名。不包括组标签。例如，如果 VarB 嵌套在列中 VarA 的下面，则可获得的变量如 CatA1_CatB1，而不是 VarA_CatA1_VarB_CatB1。

图片 21-14
从表元素构建的变量名



层变量			列变量 进	(出)	总计
高	行变量	是	1	1	2
		否	0	1	1
	总计		1	2	3
低	行变量	是	1	0	1
		否	1	1	2
	总计		2	1	3

		层变量 高 列变量 进	(出)	总计	层变量 低 列变量 进	(出)	总计
行变量	是	1	1	2	1	0	1
	否	0	1	1	1	1	2
总计		1	2	3	2	1	3

OXML 表结构

输出 XML (OXML) 是符合 spss 输出架构的 XML。有关该架构的详细说明，请参见“帮助”系统的“输出架构”部分。

- OMS 命令和子类型标识符用作 OXML 中的 **command** 和 **subType** 属性的值。示例如下：

```
<command text="Frequencies" command="Frequencies"...>

<pivotTable text="Gender" label="Gender" subType="Frequencies"...>
```

- OMS **command** 和 **subType** 属性值不受输出语言或变量名/标签或值/值标签的显示设置的影响。
- XML 区分大小写。“frequencies”的 **subType** 属性值与“Frequencies”的 **subType** 属性值不相同。
- 显示在表中的所有信息都包含在 OXML 的属性值中。在单个单元格级别，OXML 由“空”元素组成，这些元素包含属性但没有除包含在属性值中的内容之外的“内容”。
- OXML 中的表结构逐行表示；表示列的元素嵌套在行中，单个单元格嵌套在列元素中：

```
<pivotTable...>
```

```
 <dimension axis='row'...>
  <dimension axis='column'...>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
  </dimension>
 </dimension>
   ...
</pivotTable>
```

上述示例是结构的简化表示形式，显示这些元素的后代/前辈关系。但是，该示例不必显示父/子关系，因为存在通常会产生干扰作用的内嵌元素级别。

下图显示简单频率表和该表的完整的输出 XML 表示形式。

图片 21-15
简单频率表

性别

		频率	百分比	有效百分比	累积百分比
有效	女	216	45.6	45.6	45.6
	男	258	54.4	54.4	100.0
	合计	474	100.0	100.0	

图片 21-16
简单频率表的输出 XML

```
<?xml version="1.0" encoding="UTF-8" ?>
<outputTreeoutputTree xmlns="http://xml.spss.com/spss/oms"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://xml.spss.com/spss/oms
  http://xml.spss.com/spss/oms/spss-output-1.0.xsd">
 <command text="Frequencies" command="Frequencies"
  displayTableValues="label" displayOutlineValues="label"
  displayTableVariables="label" displayOutlineVariables="label">
  <pivotTable text="Gender" label="Gender" subType="Frequencies"
   varName="gender" variable="true">
   <dimension axis="row" text="Gender" label="Gender"
    varName="gender" variable="true">
    <group text="Valid">
     <group hide="true" text="Dummy">
      <category text="Female" label="Female" string="f"
       varName="gender">
       <dimension axis="column" text="Statistics">
        <category text="Frequency">
         <cell text="216" number="216"/>
        </category>
        <category text="Percent">
         <cell text="45.6" number="45.569620253165" decimals="1"/>
        </category>
```

```
      <category text="Valid Percent">
       <cell text="45.6" number="45.569620253165" decimals="1"/>
      </category>
      <category text="Cumulative Percent">
       <cell text="45.6" number="45.569620253165" decimals="1"/>
      </category>
     </dimension>
    </category>
    <category text="Male" label="Male" string="m" varName="gender">
     <dimension axis="column" text="Statistics">
      <category text="Frequency">
       <cell text="258" number="258"/>
      </category>
      <category text="Percent">
       <cell text="54.4" number="54.430379746835" decimals="1"/>
      </category>
      <category text="Valid Percent">
       <cell text="54.4" number="54.430379746835" decimals="1"/>
      </category>
      <category text="Cumulative Percent">
       <cell text="100.0" number="100" decimals="1"/>
      </category>
     </dimension>
    </category>
   </group>
   <category text="Total">
    <dimension axis="column" text="Statistics">
     <category text="Frequency">
      <cell text="474" number="474"/>
     </category>
     <category text="Percent">
      <cell text="100.0" number="100" decimals="1"/>
     </category>
     <category text="Valid Percent">
      <cell text="100.0" number="100" decimals="1"/>
     </category>
    </dimension>
   </category>
  </group>
 </dimension>
 </pivotTable>
 </command>
</outputTree>
```

您可能注意到，一个简单的小型表生成了大量的 XML。部分原因是，XML 包含一些在原始表中不明显的信息、一些甚至在原始表中不可用的信息以及一些冗余信息。

■ 显示（或应显示）在查看器的枢轴表中的表内容包含在文本属性中。示例如下：

```
<command text="Frequencies" command="Frequencies"...>
```

- 文本属性受输出语言和影响变量名/标签和值/值标签显示的设置的影响。在此示例中，**文本**属性值不相同，这取决于输出语言，而命令属性值保持相同，不管输出语言是什么。
- 只要在行标签或列标签中使用了变量或变量值，XML 都将包含一个**文本**属性以及一个或多个附加属性值。示例如下：

```
<dimension axis="row" text="Gender" label="Gender" varName="gender">
...
<category text="Female" label="Female" string="f" varName="gender">
```

- 对于数值变量，将存在**数字**属性而不是**字符串**属性。只有在变量或值具有已定义的标签时，才有**标签**属性。
- 包含数字的单元格值的 <cell> 元素将包含**文本**>属性以及一个或多个附加属性值。示例如下：

```
<cell text="45.6" number="45.569620253165" decimals="1"/>
```

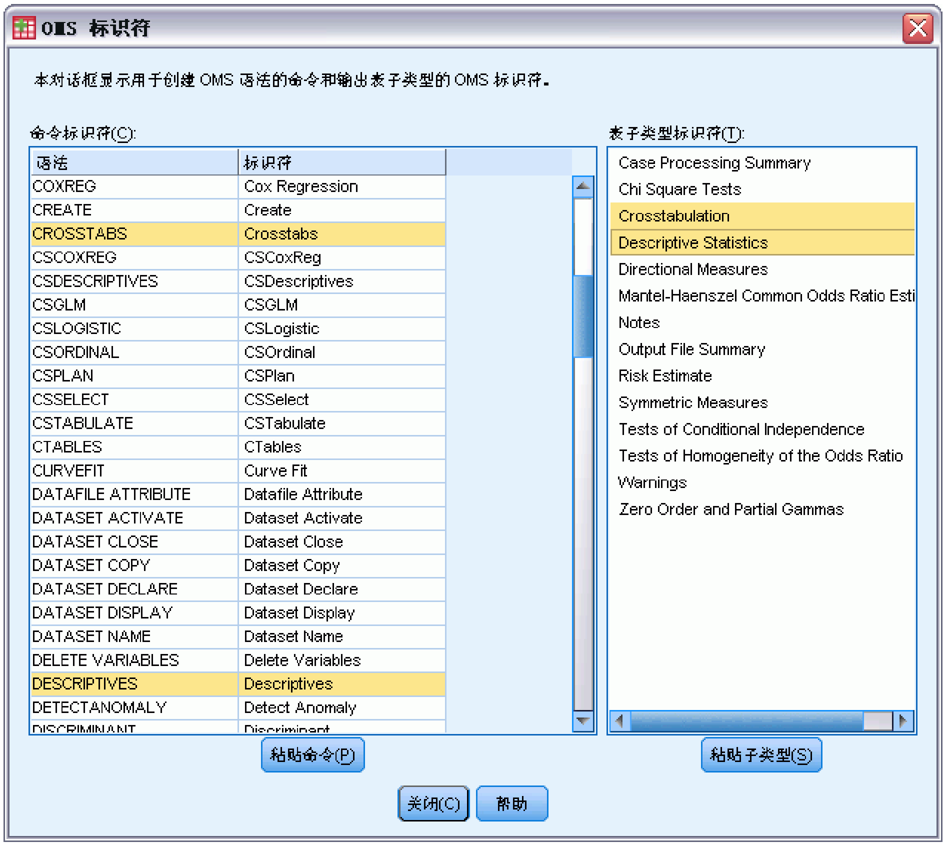
数字属性是实际的、未经过四舍五入的数值，**小数**属性指示表中显示的小数位数。

- 因为列嵌套在行中，所以标识每列的类别元素将对每行重复。例如，由于统计量显示在列中，元素 <category text="Frequency"> 在 XML 中出现了三次：一次是在男性行，一次是在女性行，一次是在总计行。

OMS 标识符

“OMS 标识符”对话框用于帮助您编写 OMS 命令语法。您可以使用此对话框将选中的命令和子类型标识符粘贴到命令语法窗口中。

图片 21-17
“OMS 标识符”对话框



使用“OMS 标识符”对话框

▶ 从菜单中选择：

实用程序 > OMS 标识符...

▶ 选择一个或多个命令标识符或子类型标识符。（按住 Ctrl 再单击可选择每个列表中的多个标识符。）

▶ 单击**粘贴命令**和/或**粘贴子类型**。

- 可用子类型的列表是根据当前选定的命令得出来的。如果选择了多个命令，可用子类型的列表就是选定的任何命令的所有可用子类型的集合。如果没有选择任何命令，则列出所有子类型。
- 标识符粘贴到指定命令语法窗口的当前光标位置。如果没有打开的命令语法窗口，则自动打开新的语法窗口。
- 每个命令和/或子类型标识符在粘贴时都用引号引起来，因为 OMS 命令语法要求使用引号。
- COMMANDS 和 SUBTYPES 关键字的标识符列表必须用方括号括起来，如：

```
/IF COMMANDS=['Crosstabs' 'Descriptives']
```

```
SUBTYPES=['Crosstabulation' 'Descriptive Statistics']
```

从查看器概要复制 OMS 标识符

您还可以从查看器概要窗格复制和粘贴 OMS 命令和子类型标识符。

▶ 在概要窗格中，右键单击项的概要条目。

▶ 选择**复制 OMS 命令标识符**或**复制 OMS 表子类型**。

这种方法与“OMS 标识符”对话框方法有一点区别：复制的标识符不会自动粘贴到命令语法窗口中。标识符只复制到剪贴板，然后可以将它粘贴到任何需要的地方。因为命令和子类型标识符的值与相应的输出 XML 格式 (OXML) 的命令和子类型的属性值相同，所以在编写 XSLT 转换时可能会觉得这种复制/粘贴方法很有用。

复制 OMS 标签

在使用 LABELS 关键字时，您可以复制标签（而不是标识符）。标签可以用于区分同一类型的多个图形或表格，在这些图形或表格中，概要文本反映特定输出对象（如变量名或标签）的某个属性。但是有一些因素会影响标签文本：

- 如果正在处理拆分文件，则拆分文件组标识将会附加到标签上。
- 包括关于变量或值的信息的标签受概要窗格（“编辑”菜单，“选项”，“输出标签”选项卡）中的变量名/标签和值/值标签的显示设置的影响。
- 标签受当前输出语言设置（“编辑”菜单，“选项”，“常规”选项卡）的影响。

复制 OMS 标签

▶ 在概要窗格中，右键单击项的概要条目。

▶ 选择**复制 OMS 标签**。

与命令标识符和子类型标识符一样，标签必须用引号引起来，整个列表必须用方括号括起来，如：

```
/IF LABELS=['Employment Category' 'Education Level']
```

章

22

脚本编写工具

脚本编写工具使您可以将一些任务自动化，包括：

- 打开和保存数据文件。
- 以多种格式将图表导出为图形文件。
- 在“查看器”中定制输出。

脚本语言

可用脚本语言取决于使用的平台。对于 Windows 平台，可用脚本语言为 Basic（随 Core 系统一起安装）和 Python 编程语言。对于所有其他平台，脚本以 Python 编程语言提供。

要启用使用 Python 编程语言编写脚本的功能，必须安装 Python 和 IBM® SPSS® Statistics - Integration Plug-In for Python。 有关信息，请参阅帮助系统的“Core 系统>常见问题解答”中的“如何获得集成插件”。

注意：SPSS Inc. 并不是 Python 软件的所有者或许可方。任何 Python 用户必须同意 Python 网站上 Python 许可协议的条款。SPSS Inc. 不对 Python 程序的质量作出任何形式的声明。SPSS Inc. 拒绝承担与 Python 程序使用相关的全部责任。

默认脚本语言

默认脚本语言确定创建新脚本时启动的脚本编辑器。它还指定了其可执行程序将用于运行自动脚本的默认语言。在 Windows 中，默认脚本语言为 Basic 语言。可以从“选项”对话框中的“脚本”选项卡中更改默认语言。

样本脚本

在安装 IBM® SPSS® Statistics 的目录的 Samples 子目录中包含一些脚本。可以不做任何更改使用这些脚本，或者可以按照您的需要进行定制。

创建新脚本

▶ 从菜单中选择：

文件 > 新建 > 脚本

与默认脚本语言相关的编辑器打开。

运行脚本

▶ 从菜单中选择：

实用程序 > 运行脚本...

▶ 选择需要的脚本。

▶ 单击**运行**。

除了从实用程序>运行脚本中运行以外，Python 脚本还可用多种方式运行。

编辑脚本

▶ 从菜单中选择：

文件 > 打开 > 脚本...

▶ 选择需要的脚本。

▶ 单击**打开**。

在与该脚本编写所用语言相关的编辑器中打开该脚本。

自动脚本

自动脚本即当被从已选过程中创建的特定输出触发时自动运行脚本。例如，可以使用自动脚本在“双变量相关”过程生成“相关”表时自动移去上对角线并突出显示低于某个显著水平的相关系数。

自动脚本可以特定于一个给定过程和输出类型或应用于不同过程中的特定输出类型。例如，您可能拥有一个既能为单因素 ANOVA 生成的 ANOVA 表设置格式，又能为由其他统计过程生成的 ANOVA 表设置格式的脚本。另一方面，“频率”过程生成频率表和统计量表，而您可以选择两个表各自具有不同的自动脚本。

给定过程的每个输出类型只可与一个单一自动脚本相关。但是仍可在应用任何特定输出类型的自动脚本之前创建可应用于所有新建“查看器”项的基础自动脚本。

“选项”对话框中的“脚本”选项卡（从“编辑”菜单进入）显示您系统中配置的自动脚本并使您能创建新的自动脚本或修改已有自动脚本的设置。或者，还可以从“查看器”中直接创建并配置输出项的自动脚本。

触发自动脚本的事件

以下事件可以触发自动脚本：

- 枢轴表的创建
- “注释”对象的创建
- “警告”对象的创建

也可以使用脚本间接触发一个自动脚本。例如，可以编写调用 Correlations 过程的脚本，而 Correlations 过程反过来触发注册到产生的 Correlations 表中的自动脚本。

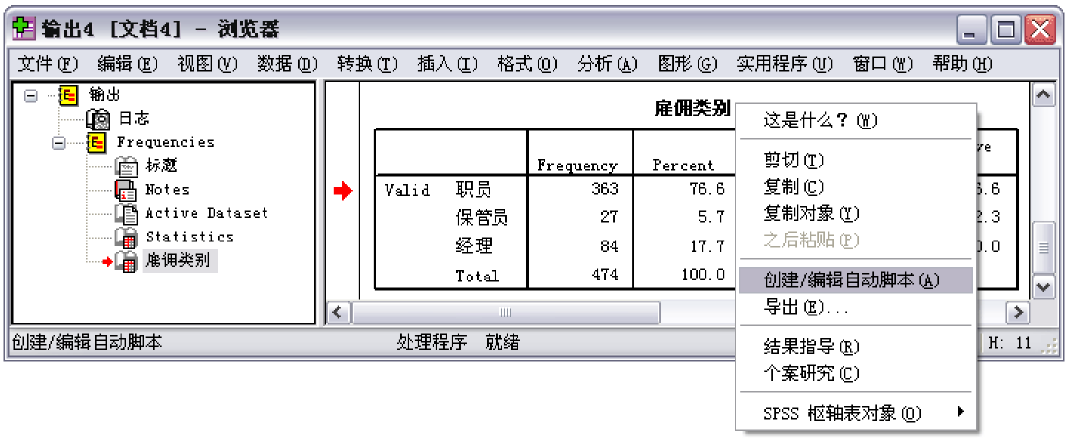
创建自动脚本

可以从要用作触发器的输出对象开始创建自动脚本 — 例如，频率表。

▶ 在“查看器”中，选择将要触发自动脚本的对象。

▶ 从菜单中选择：
实用程序 > 创建/编辑自动脚本…

图片 22-1
创建新的自动脚本



若已选对象并不具有关联自动脚本，则“打开”对话框将提示您输入新脚本的位置与名称。

▶ 浏览至新脚本将被保存的位置，输入文件名并单击**打开**。默认脚本语言编辑器打开。可以从“选项”对话框中的“脚本”选项卡中更改默认脚本。

▶ 键入代码。

有关转换在 16.0 以前版本中使用的定制 Sax Basic 自动脚本的帮助，请参见与 16.0 之前版本的兼容性第 377 页码。

注意：默认情况下，与默认脚本语言相关的可执行文件将被用于自动脚本的运行。可以从“选项”对话框中的“脚本”选项卡中更改可执行文件。

若已选对象已与自动脚本关联，则该脚本在脚本编辑器中以与之相关的脚本编写语言打开。

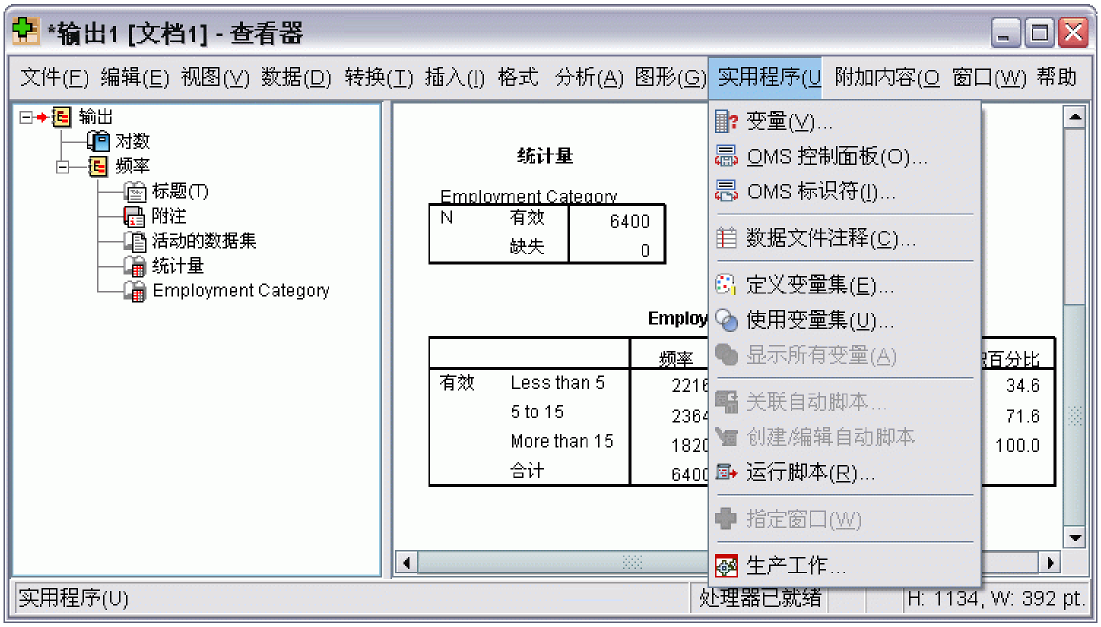
关联已有脚本与查看器对象

可以通过将其与“查看器”中的已选对象关联 — 例如，频率表，将已有脚本用作自动脚本。

▶ 在“查看器”中，选择一个对象与自动脚本关联（多个“查看器”对象可以触发同一自动脚本，但每个对象只能与单一自动脚本关联）。

▶ 从菜单中选择：
实用程序 > 关联自动脚本…

图片 22-2
关联自动脚本



若已选对象并不具有关联自动脚本，则“选择自动脚本”对话框打开。

▶ 浏览所需脚本并选中。

▶ 单击**应用**。

若已选对象已与自动脚本关联，则提示您对所需更改的关联进行验证。单击**确定**打开“选择自动脚本”对话框。

或者，还可以在“选项”对话框中的“脚本”选项卡中将已有脚本配置为自动脚本。自动脚本可被应用于已选输出类型集或特定为应用于所有新建“查看器”项的基础自动脚本。

以 Python 编程语言编写脚本

IBM® SPSS® Statistics 在 Windows、Linux 和 Mac OS 中为以 Python 语言编程提供两个单独界面。这些界面的使用需要 IBM® SPSS® Statistics - Essentials for Python 随附的 IBM® SPSS® Statistics - Integration Plug-In for Python。 有关信息，请参阅帮助系统的“Core 系统>常见问题解答”中的“如何获得集成插件”。有关开始使用 Python 编程语言的帮助，请参见 Python 教程，从以下地方获取 http://docs.python.org/tut/tut.html

Python 脚本

Python 脚本利用 Python **SpssClient** 模块提供的界面。它们在用户界面和输出对象上操作并能运行命令语法。例如，可以使用 Python 脚本自定义枢轴表。

- Python 脚本可以从实用程序>运行脚本或从 Python 编辑器中运行，该编辑器可以从 SPSS Statistics（从“文件>打开>脚本”进入）中启动，或从外部 Python 过程，如 Python IDE 或 Python 翻译器中运行。
- Python 脚本可作为自动脚本运行。
- Python 脚本在 SPSS Statistics 客户端运行的机器上运行。

Python 脚本的可用 SPSS Statistics 类和方法的完整文档可以在《IBM SPSS Statistics 脚本编写指南》中找到，它位于帮助系统中 Integration Plug-In for Python 的下面，安装 Essentials for Python 后便可使用。

Python 程序

Python 程序利用 Python **spss** 模块提供的界面。它们在 SPSS Statistics 处理器上操作并被用于控制命令语法工作流、读取及写入活动数据集、创建新数据集及创建能生成其自有枢轴表输出的客户过程。

- Python 程序可从命令语法中与 **BEGIN PROGRAM-END PROGRAM** 块共同运行，或从外部 Python 过程，如 Python IDE 或 Python 翻译器中运行。
- Python 程序不可作为自动脚本运行。
- 在分布式分析模式下（可用于 SPSS Statistics 服务器），Python 程序在 SPSS Statistics 服务器运行的电脑上执行。

有关 Python 程序的更多信息，包括其提供的 SPSS Statistics 函数和类的完整文档，请参见 IBM SPSS Statistics 的 Python 集成程序包的文档说明，它位于帮助系统中 Integration Plug-In for Python 的下面，安装 Essentials for Python 后便可使用。

运行 Python 脚本与 Python 程序

Python 脚本与 Python 程序均可与 IBM® SPSS® Statistics 共同运行，或从外部 Python 过程，如 Python IDE 或 Python 翻译器中运行。

Python 脚本

从 SPSS Statistics 中运行“Python 脚本”。可以从实用程序>运行脚本中运行 Python 脚本或从 Python 脚本编辑器中运行，该编辑器在从文件>打开>脚本中打开 Python 文件（.py）时启动。从 Python 编辑器中运行的脚本在启动编辑器的 SPSS Statistics 客户端上操作，该编辑器从 SPSS Statistics 中启动。这使您能从 Python 编辑器中调试您的 Python 代码。

从“外部 Python 过程”中运行 Python 脚本。 可以从任何外部 Python 过程中运行 Python 脚本，例如不从 SPSS Statistics 中启动的 Python IDE 或 Python 翻译器。脚本将尝试连接到现有 SPSS Statistics 客户端。如果找到多个客户端，将连接到最近启动的客户端。如果现有客户端未找到，Python 脚本启动 SPSS Statistics 客户端的新实例。默认状态下，“数据编辑器”与“查看器”在新客户端中均不可视。可以选择使其可视或在不可视模式下使用数据集与输出文档。

Python 程序

从“命令语法”中运行 Python 程序。 可以通过在命令语法中向 BEGIN PROGRAM-END PROGRAM 块中嵌入 Python 代码运行 Python 程序。可以从 SPSS Statistics 客户端中或 SPSS Statistics 批处理工具 — 即由 SPSS Statistics Server 提供的一个单独可执行文件中运行命令语法。

从“外部 Python 过程”中运行 Python 程序。 可以从任何外部 Python 过程中运行 Python 程序，例如 Python IDE 或 Python 翻译器。在此模式下，Python 程序无需 SPSS Statistics 客户端的相关实例即可启动 SPSS Statistics 处理器的新实例。可以通过所选的 Python IDE，使用此模式调试您的 Python 程序。

从“Python 程序”中调用“Python 脚本”，反之亦然

从“Python 程序”中运行“Python 脚本”。 您可以通过导入包含脚本的 Python 模块并调用执行脚本的模块中的函数从 Python 程序中运行 Python 脚本。您还可直接从 Python 程序中调用 Python 脚本方法。 在从外部 Python 进程运行 Python 程序，或从 SPSS Statistics 批处理工具（随 SPSS Statistics 服务器提供）运行 Python 程序时，这些功能不可用。

从“Python 程序”中触发的“Python 自动脚本”。 特定为自动脚本的 Python 脚本将在 Python 程序执行过程时被触发，该过程包含与自动脚本关联的输出项。例如，将一个自动脚本与由“描述”过程生成的“可描述统计数据”表关联。然后运行执行“描述”过程的 Python 程序。Python 自动脚本将被执行。

从“Python 脚本”中运行“Python 程序”。 Python 脚本可以运行命令语法，这就意味着它们能运行包含 Python 程序的命令语法。

限制与警告

- 从由 SPSS Statistics 启动的 Python 编辑器中运行 Python 程序将启动 SPSS Statistics 处理器的新实例且不与由编辑器启动的 SPSS Statistics 实例产生交互作用。
- Python 程序不被用于从实用程序>运行脚本中运行。
- Python 程序不可作为自动脚本运行。
- 由 spss 模块提供的界面无法在 Python 脚本中使用。

“Python 编程语言”的“脚本编辑器”

对于 Python 编程语言而言，默认编辑器为 Python 提供的 IDLE。IDLE 提供有限功能集的集成开发环境（IDE）。许多 IDE’s 均适用于 Python 编程语言。例如，在 Windows 中，可以自由选择使用 PythonWin IDE。

更改 Python 编程语言的脚本编辑器：

▶ 打开位于其安装 IBM® SPSS® Statistics 目录中的文件 clientscriptingcfg.ini。注意：必须使用支持 UTF-16 的文本编辑器来编辑 clientscriptingcfg.ini，例如 Windows 上的 SciTE，或 Mac 上的 TextEdit 应用程序。

▶ 在标有 [Python] 的部分下，更改 EDITOR_PATH 值以指出所需编辑器的可执行文件。

- 同一部分中，更改 EDITOR_ARGS 值以处理需传递至编辑器的任何参数。若无需任何参数，移除一切已有值。

Basic 语言中的脚本编写

Basic 语言中的脚本编写只适用于 Windows 并以 Core 系统安装。关于 Basic 语言中脚本编写的拓展在线帮助可从 IBM® SPSS® Statistics “Basic 语言脚本编辑器”获取。
当默认脚本语言（从“选项”对话框的“脚本”选项卡中设置）被设置为 Basic 语言（Windows 的系统默认语言）时，可以从文件>新建>脚本中进入编辑器。还可通过在类型列表的“文件”中选择 Basic (wwd;sbs)，从文件>打开>脚本中进入。

与 16.0 之前版本的兼容性

废弃方法与属性

一些自动化方法与属性已被版本 16.0 及以上版本废弃。就常规功能而言，它包括了所有与互动图像相关的对象、“草稿文档”对象及与地图相关的方法与属性。更多详情，参见“IBM® SPSS® Statistics Basic 语言脚本编辑器”所提供帮助系统中的“16.0 版本的发布注释”。从脚本编辑器中的帮助>SPSS Statistics 对象帮助中进入 SPSS Statistics 特定帮助。

全局过程

16.0 之前的版本中，脚本编写工具包含全局过程文件。即使 16.0 之前的 Global.sbs（更名为 Global.wwd）版本中安装了向后兼容性，16.0 及以上版本中，脚本编写工具并未使用全局过程文件。

要迁移全局过程文件中名为功能的 16.0 之前版本脚本，将语句 '#Uses "<install dir>\Samples\Global.wwd" 添加至其安装 <install dir> 目录 SPSS Statistics 的脚本申报部分。'#Uses 为 Basic 语言脚本处理器识别的特殊注释。若不确定脚本是否使用了全局过程文件，则应添加 '#Uses 语句。也可使用 '$Include:代替 '#Uses。

Legacy 自动脚本

16.0 之前的版本中，脚本编写工具包括包含了所有自动脚本的单一自动脚本文件。16.0 及以上版本则无单一自动脚本文件。现在每个自动脚本均被保存在单独文件中并可应用于一个或多个输出项中，相反在 16.0 之前版本中每个自动脚本均特定于一个个别输出项。
　随 16.0 以前版本安装的一些自动脚本位于 SPSS Statistics 安装目录的 Samples 子目录中，作为一组单独的脚本文件提供。它们由以 Autoscript 结尾的文件名和 wwd 的文件类型标识。默认情况下，它们与任何输出项均无关联。从“选项”对话框的“脚本”选项卡中进行关联。

任何用于 16.0 之前版本的定制自动脚本必须手动转换并从“选项”对话框的“脚本”选项卡中关联至一个或多个输出项。转换过程包括以下步骤：

- 从 legacy Autoscript.sbs 文件中提取特定自动脚本的子例程并以扩展名 wwd 或 sbs 将其存为新建文件。文件名随意。

- 将子例集名称改为 Main 并删除参数规范，保持对脚本所需参数，如触发自动脚本的枢轴表的追踪。
- 使用 scriptContext 对象（总是可用）获取自动脚本所需值，如触发自动脚本的输出项。
- 从“选项”对话框的“脚本”选项卡中将脚本文件关联至所需输出对象。

为说明转换代码，请考虑来自 legacy Autoscript.sbs 文件的自动脚本 Descriptives_Table_DescriptiveStatistics_Create。

```
Sub Descriptives_Table_DescriptiveStatistics_Create _
(objPivotTable As Object,objOutputDoc As Object,lngIndex As Long)
“自动脚本
”触发事件：运行可描述性过程后，描述性
'          统计数据表的创建。

“目的：将当前活动枢轴表中的“行”与“列”对调。
“假设：已选“枢轴表”已激活。
“影响：将输出中的“行”与“列”对调
“输入：“枢轴表”、OutputDoc、项索引

   Dim objPivotManager As ISpssPivotMgr
   Set objPivotManager=objPivotTable.PivotManager
   objPivotManager.TransposeRowsWithColumns

End Sub
```

以下为已转换脚本：

```
Sub Main

'Purpose:将当前活动枢轴表中的“行”与“列”对调。
“影响：将输出中的“行”与“列”对调

   Dim objOutputItem As ISpssItem
   Dim objPivotTable as PivotTable
   Set objOutputItem = scriptContext.GetOutputItem()
   Set objPivotTable = objOutputItem.ActivateTable

   Dim objPivotManager As ISpssPivotMgr
   Set objPivotManager = objPivotTable.PivotManager
   objPivotManager.TransposeRowsWithColumns
   objOutputItem.Deactivate
End Sub
```

- 注意，已转换脚本中并未指明该脚本会应用哪个对象。输出项与自动脚本间的关联从“选项”对话框的“脚本”选项卡中设置并在会话之间保持不变。
- scriptContext.GetOutputItem 获得触发自动脚本的输出项（一个 ISpssItem 对象）。
- 以 scriptContext.GetOutputItem 返回的对象未被激活。若您的脚本需要激活对象，则需要将其激活，如本示例中所示的 ActivateTable 方法。当一切表操作完成，调用**取消激活**方法。

对于版本 16.0，在作为自动脚本运行的脚本与未作为自动脚本运行的脚本间并无差别。任何进行过相应编码的脚本均可用于所有上下文中。

注意：要从应用程序创建事件触发脚本，请参见启动脚本。

脚本编辑器

对于 16.0 及以上版本，用于 Basic 语言的脚本编辑器不再支持以下 16.0 之前功能：

- 脚本、分析、图形、实用程序、插件菜单。
- 将命令语法粘贴至脚本窗口的能力。

“SPSS Statistics Basic 语言脚本编辑器”是通过文件>新建>脚本、文件>打开>脚本或实用程序>创建/编辑自动脚本（从“查看器”窗口中）从 SPSS Statistics 中启动的独立应用程序。使您能针对从中启动的 SPSS Statistics 实例运行脚本。已经打开，编辑器将一直开启直至退出 SPSS Statistics，但是用 SPSS Statistics 对象的脚本将不再运行。

文件类型

对于 16.0 及以上版本，脚本编写工具将继续支持 sbs 文件类型的脚本运行与编辑。默认状态下，通过“SPSS Statistics Basic 语言脚本编辑器”创建的新 Basic 语言脚本的文件类型为 wwd。

使用“外部 COM 客户端”

对于 16.0 及以上版本，用于例示“外部 COM 客户端”中 SPSS Statistics 的程序识别器为 SPSS.Application16。应用程序对象应声明为 spsswinLib.Application16。例如：

```
Dim objSpssApp As spsswinLib.Application16
Set objSpssApp=CreateObject("SPSS.Application16")
```

要连接到一个来自外部 COM 客户端的 SPSS Statistics 客户端运行实例，请使用：

```
Dim objSpssApp As spsswinLib.Application16
Set objSpssApp=GetObject("","SPSS.Application16")
```

如果多个客户端正在运行，GetObject 将连接到最近启动的客户端。

注意：对于 16.0 之后的版本，标识仍然是 Application16。

脚本上下文对象

脚本作为自动脚本运行时的检测

利用 scriptContext 对象，可以在脚本作为自动脚本运行时进行检测。这使您可以对脚本进行编码以便其在所有上下文中可以运行（无论自动脚本与否）。此普通脚本说明了方法。

```
Sub Main
   If scriptContext Is Nothing Then
    MsgBox "I'm not an autoscript"
   Else
    MsgBox "I'm an autoscript"
   End If
End Sub
```

- 当脚本不作为自动脚本运行时，scriptContext 对象将具有无值。
- 本示例中被赋予 If-Else 逻辑，应在 Else 条款中包含自动脚本特定代码。任何不在自动脚本上下文中运行的代码均应包含在 If 条款中。当然也可以包含将要在所有上下文中运行的代码。

获区自动脚本所需值

scriptContext 对象提供自动脚本所需值的使用，如触发当前自动脚本的输出项。

- scriptContext.GetOutputItem 方法返回至触发当前自动脚本的输出项（一个 ISpssItem 对象）。
- scriptContext.GetOutputDoc 方法返回至与当前自动脚本关联的输出文档（一个 ISpssOutputDoc 对象）。
- scriptContext.GetOutputItemIndex 方法返回至相关输出文档中触发当前自动脚本的输出项索引。

注意：以 scriptContext.GetOutputItem 返回的对象未被激活。若您的脚本需要激活对象，则需要将其激活 — 例如通过 ActivateTable 方法。当一切操作完成，调用**取消激活**方法。

启动脚本

您可以创建在每次会话启动时运行的脚本，以及在每次切换服务器时运行的单独脚本。在 Windows 上，您可以使用这些脚本的 Python 和 Basic 版本。对于所有其他平台，只能使用 Python 脚本。

- 启动脚本必须命名为 StartClient_.py(Python) 或 StartClient_.wwd(Basic)。
- 在切换服务器时运行的脚本必须命名为 StartServer_.py(Python) 或 StartServer_.wwd(Basic)。
- 脚本必须位于安装目录下的 scripts 目录中。对于 Windows 和 Linux，此目录在安装目录的根目录下；对于 Mac，此目录在应用程序包中的 Contents 目录下。注意，不论您是否在分布模式下工作，所有脚本（包括 StartServer_ 脚本）必须位于客户端计算机上。
- 在 Windows 上，如果 scripts 目录同时包含 Python 和 Basic 版本的 StartClient_ 或 StartServer_，则两种版本的脚本都会执行。执行顺序为 Python 版本优先于 Basic 版本。
- 如果您的系统被配置成以分布式模式启动，则当每次会话启动时，先运行 StartClient_ 脚本，然后再运行 StartServer_ 脚本。注意：StartServer_ 脚本还会在每次切换服务器时运行，但 StartClient_ 脚本仅在会话启动时运行。

附录

A

TABLES 和 IGRAPH 命令语法转换器

如果您的命令语法文件包含 TABLES 语法，您要将其转换为 CTABLES 语法和/或您要将 IGRAPH 语法转换为 GGRAPH 语法，可以使用一个简单实用程序进行转换。但是，TABLES 与 CTABLES 以及 IGRAPH 与 GGRAPH 在功能上有很大差别。您可能会发现该实用程序无法转换某些 TABLES 和 IGRAPH 语法作业，或者转换得到的 CTABLES 和 GGRAPH 语法所生成的表格和图形可能与 TABLES 及 IGRAPH 命令所生成的原始表格和图形差别较大。对于大多数的表，您可以编辑转换后的语法，使之生成与原始表十分相似的表。

该实用程序设计用于：

- 根据现有语法文件创建新的语法文件。不更改原始语法文件。
- 只转换语法文件中的 TABLES 及 IGRAPH 命令。不更改该文件中的其他命令。
- 以注释形式保留原始 TABLES 及 IGRAPH 语法。
- 用注释标识每个转换块的开头和结尾。
- 标识无法转换的 TABLES 和 IGRAPH 语法命令。
- 转换符合交互式或生产模式语法规则的命令语法文件。

此实用程序无法转换包含错误的命令。还具有以下其他限制。

TABLES 限制

在某些情况下，该实用程序对 TABLES 命令的转换可能不正确，例如具有以下特点的 TABLES 命令：

- TABLES 子命令中包含以“sta”或“lab”开头的带括号的变量名（如果变量整体放在括号中）；例如，var1 by (statvar) by (labvar)。这些变量将被解释为 (STATISTICS) 和 (LABELS) 关键字。
- 包含的 SORT 子命令使用缩写字母 A 或 D 指示升序或降序。这些缩写字母将被解释为变量名。

该实用程序无法转换具有以下特点的 TABLES 命令：

- 包含语法错误。
- 包含使用 TO 关键字引用变量范围的 OBSERVATION 子命令（例如，var01 TO var05）。
- 包含用加号分隔成多部分的字符串（例如，TITLE "My" + "Title"）。
- 包含在没有宏扩展时会导致 TABLES 语法无效的宏调用。由于转换器不扩展宏调用，因此会将宏调用简单地视为标准 TABLES 语法的一部分。

该实用程序不转换包含在宏中的 TABLES 命令。所有宏都不受转换过程的影响。

IGRAPH 限制

IGRAPH 在 16 版本中更改较大。由于这些更改，在该版本之前的 IGRAPH 语法中创建的某些子命令和关键字可能不被接受。有关完整修正历史，请参见 IGRAPH 一节（命令语法索引）。

转换实用程序可能会在 GGRAPH 语法中生成存储在 INLINETEMPLATE 关键字中的其他语法。此关键字只能通过转换程序创建。用户无法编辑其语法。

使用转换实用程序

转换实用程序 SyntaxConverter.exe 可以在安装目录中找到。该实用程序设计为从命令提示符处运行。命令的一般形式为：

```
syntaxconverter.exe [path]/inputfilename.sps [path]/outputfilename.sps
```

您必须从安装目录运行此命令。

如果有任何目录名中包含空格，应当用引号将路径和文件名整体括起来，如下所示：

```
syntaxconverter.exe /myfiles/oldfile.sps "/new files/newfile.sps"
```

交互式与生产模式命令语法规则

转换实用程序可以转换使用交互式或生产模式语法规则的命令文件。

交互式。交互式语法规则为：

- 每条命令都另起一行。
- 每条命令都以句点 (.) 结尾。

生产模式。 生产工具和文件中的命令（通过在其他命令文件中使用 INCLUDE 命令访问）使用生产模式语法规则：

- 每条命令必须从新的一行的第一列开始。
- 连续行至少须缩进一个空格。
- 命令结尾处的句点是可选的。

如果命令文件使用生产模式语法规则，并且每条命令都不以句点结尾，则在运行 SyntaxConverter.exe 时需要包含命令行开关 -b（或 /b），如下所示：

```
syntaxconverter.exe -b /myfiles/oldfile.sps /myfiles/newfile.sps
```

SyntaxConverter 脚本（仅 Windows）

在 Windows 上，您也可以通过脚本 SyntaxConverter.wwd 运行语法转换器，该脚本位于安装目录中的 Samples 目录中。

▶ 从菜单中选择：
实用程序 > 运行脚本...

▶ 浏览至 Samples 目录并选择 SyntaxConverter.wwd。

这将打开一个简单的对话框，在该对话框中可以指定新旧命令语法文件的名称和位置。

附录

B

Notices

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Trademarks

索引